THE PHYSICAL REVIEW

A journal of experimental and theoretical physics established by E. L. Nichols in 1893

Published by the American Institute of Physics for the

AMERICAN PHYSICAL SOCIETY

Officers

President

E. T. Bowen
University of California
Berkeley, California

Vice-President

E. P. Wigner
Princeton University
Princeton, New Jersey

Secretary

K. K. Darrow
Bell Telephone Laboratories
Office:
Columbia University
New York 27, New York

Treasurer

G. B. Pegram
Columbia University
New York 27, New York

Local Secretary
for the Pacific Coast

William A. Nierenberg
University of California
Berkeley, California

Managing Editor

S. A. Goudsmit
Brookhaven National Laboratory
Upton, Long Island
New York

BOARD OF EDITORS

Editor

S. A. Goudsmit

Assistant Editor

S. Pasternack

Associate Editors

1953–1955	1954–1956	1955–1957
W. A. Fowler	P. G. Bergmann	G. J. Dienes
Kenneth Greisen	Gregory Breit	R. G. Herb
Charles Kittel	F. J. Dyson	K. G. McKay
R. E. Marshak	P. Kusch	J. R. Pellam
W. K. H. Panofsky	R. Smoluchowski	Lewi Tonks
V. F. Weisskopf	J. C. Street	Gregor Wentzel

Manuscripts for publication should be submitted to S. A. Goudsmit, Brookhaven National Laboratory, Upton, Long Island, New York. The authors' institutions are requested to pay a publication charge of $25 per page which, if honored, entitles them to 100 free reprints. Instructions will be sent with galley proofs.

Proof and all correspondence concerning papers in the process of publication should be addressed to the Publication Manager, American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Rates

United States and Canada $30.00
Elsewhere $32.00

Rates to Members[1]

Yearly back number rate, when available, prior to distribution
Prior to 1949: $25.00
Thereafter: $33.00

Single copies: $3.00 each, prior to July, 1939;
$1.50 each, July, 1939–December, 1941;
$3.00 each, January, 1942–December, 1944;
$1.50 each, thereafter.

General index, 1893–1930: $4.00.

1931–1950: Paper binding, $10 (Members); $14 (Nonmembers)
Cloth binding, $12 (Members); $16 (Nonmembers)

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Changes of address in the case of members of the American Physical Society should be addressed to the Treasurer; in the case of other subscribers to the American Institute of Physics.

[1]For Series I inquire of Physics Department, Cornell University, Ithaca, New York.

The Physical Review is published semi-monthly at Prince and Lemon Streets, Lancaster, Pennsylvania.

Entered as second-class matter March 6, 1918, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.

Acceptance for mailing at special rate of postage provided for in paragraph (d-2), section 34.40, P.L. & R. of 1948, authorized January 22, 1929.

THE
PHYSICAL REVIEW

A journal of experimental and theoretical physics established by E. L. Nichols in 1893

SECOND SERIES, VOL. 100, No. 2 OCTOBER 15, 1955

Particle Transport, Electric Currents, and Pressure Balance in a Magnetically Immobilized Plasma

P. STEHLE
University of Pittsburgh, Pittsburgh, Pennsylvania
(Received June 27, 1955)

The expression for the pressure gradient in a magnetically immobilized plasma,

$$\nabla p = -(1/8\pi)\nabla(B^2)$$

is derived without detailed study of particle orbits and still showing the various effects which appear in a detailed treatment.

IN a paper with the above title, Tonks[1] has discussed the pressure balance equation in a magnetically confined plasma from a completely microscopic viewpoint. This result is the same as that of Spitzer[2] who also derived the equation by an averaging process over individual trajectories. It is illuminating to see the result arise without treating orbits in any detail.

Consider only Spitzer's (1,0) approximation (first-order effects in field inhomogeneities, no collision effects). A plasma is located in a homogeneous external field $H_x = H_0$. Values of physical quantities depend only on x. Ions and electrons spiral about the field lines, giving rise to a current. The ion current due to motion in a circle about the field can be described by a density of magnetic moment, M_i, given by

$$M_i = -\left\langle \frac{n_i e v_{i\perp}}{2\pi r c} \pi r^2 \right\rangle = -\frac{n_i m_i \langle v_{i\perp}^2 \rangle}{2B}, \tag{1}$$

where n_i is the ion density, m_i the ion mass, $v_{i\perp}$ the speed perpendicular to the field, and B is the local field. Defining the kinetic temperature of the ions by $kT = m_i v_{i\perp}^2/2$, and assuming the kinetic temperature of the electrons to be the same for simplicity, we get for the magnetic moment density due to both ions and electrons

$$M = -p/B, \tag{2}$$

where $p = nkT$ is the pressure of the plasma.

The current due to the drift of ions across the field when the field is inhomogeneous is in the y-direction and is[3]

$$v_{iD} = m_i v_{i\perp}^2 c B'/2B^2 e_i, \tag{3}$$

so that the net current density due to ions and electrons is

$$j = pcB'/B^2 = -(c/4\pi)H_1', \tag{4}$$

where H_1 is the magnetic field due to this current distribution.

Finally, then:

$$\begin{aligned} B &= H + 4\pi M \\ &= H_0 + H_1 - (4\pi p/B), \end{aligned} \tag{5}$$

so that

$$B' = H_1' + (4\pi p B'/B^2) - (4\pi p'/B). \tag{6}$$

The first term, representing the drift due to field inhomogeneity, just compensates for the second term representing, as Tonks has shown, the effect of the change in orbital radius coming from the field inhomogeneity, leaving only the density gradient term. Thus

$$p' = (-1/8\pi)(B^2)'. \tag{7}$$

The virtue of this approach lies in the avoidance of detailed consideration of orbits.

This work results from discussions with Dr. G. Gibson and Dr. J. L. Johnson.

[1] L. Tonks, Phys. Rev. **97**, 1443 (1955).
[2] L. Spitzer, Astrophys. J. **116**, 299 (1952).
[3] H. Alfven, *Cosmical Electrodynamics* (Oxford University Press, New York, 1950), p. 22.

Fluctuations on a Damped Transmission Line

C. W. McCombie
Department of Natural Philosophy, University of Aberdeen, Aberdeen, Scotland
(Received March 22, 1955; revised manuscript received July 5, 1955)

It is shown that, contrary to what has been stated, a straightforward application of the equipartition theorem to electrical fluctuations on a damped transmission line leads to results in agreement with those obtained from Nyquist's theorem.

I. INTRODUCTION

TWO familiar procedures are available for establishing the mean fluctuation energy in a mode of vibration (standing wave) on a transmission line at temperature T. An appeal to the equipartition theorem leads at once to the value kT for this energy: quantum corrections, important only for very high frequency modes, are ignored throughout the present paper. The alternative procedure is to treat the fluctuations as arising from the Nyquist voltages associated with the various elements of resistance on the line. Rhodes[1] has made a somewhat indirect application of the second procedure and has concluded that the mean fluctuation energy of a mode has a value which depends on the amount of damping on the line and so, in general, differs from kT.

Both the procedures mentioned involve considerable assumptions.[2] Except in special cases[3,4] these have not, to our knowledge, been shown to be justified by the basic principles of statistical mechanics or kinetic theory. Thus the application made here of the equipartition theorem rests on the assumption that, in using statistical mechanical methods to find the mean square value of a macroscopic coordinate, we may ignore those interactions which give rise macroscopically only to damping forces. Nyquist calculations, on the other hand, depend on the assumption that we may associate with a seat of damping a fluctuating force with statistical properties independent of the system in which the seat of damping is included. This last assumption is required before we can attach any general significance to the statistical properties determined for the Nyquist forces by requiring that they reproduce statistical mechanical results for some special system (an undamped transmission line terminated by matched resistances in Nyquist's original treatment).

Even in the absence of precise theoretical justification, however, we have reason to have considerable confidence in the two procedures. They have been applied in a straightforward way to a large class of linear electrical and dynamical systems, and have always agreed as to their predictions. These predictions have, moreover, been borne out by the available experimental results. A disagreement between the two procedures in the case of the damped transmission line would therefore be surprising and interesting. It will, however, be shown in what follows that the two methods do in fact lead to consistent results when applied in a straightforward way to a continuous transmission line. In particular, the Nyquist procedure does not predict a dependence of the mean fluctuation energy on the amount of damping.

Detailed consideration will be given to a transmission line of arbitrary length L with its ends shorted. A very similar discussion can be given for a line with open ends. The results obtained will be shown to hold for a very long line terminated in an arbitrary way: this is the case considered by Rhodes.

The line will be supposed to have resistance r, inductance l, leakage conductance ρ, and capacity c per unit length. The solutions with sinusoidal dependence on time of the equations satisfied by the voltage V and current I will, on an unperturbed length of line, have the space dependence

$$V=Ae^{\gamma y}+Be^{-\gamma y}, \quad I=-(1/Z)(Ae^{\gamma y}-Be^{-\gamma y}), \quad (1)$$

where y is measured along the line, A and B are disposable constants, and

$$\gamma^2=(r+2\pi i\nu l)(\rho+2\pi i\nu c),$$
$$Z^2=(r+2\pi i\nu l)/(\rho+2\pi i\nu c). \quad (2)$$

This specifies our notation.

II. STATISTICAL MECHANICS OF MACROSCOPIC COORDINATES

In applying the statistical mechanics of macroscopic coordinates to the present problem, one introduces coordinates which describe the macroscopic electrical properties of the line and which in the absence of damping would be normal coordinates. For a line with shorted ends, these coordinates are the coefficients in a Fourier sine analysis of the voltage and in a Fourier cosine analysis of the current, i.e., the a's and b's in

$$V=\sum_{s=1}^{\infty} a_s \sin(\pi sy/L), \quad I=\sum_{s=0}^{\infty} b_s \cos(\pi sy/L). \quad (3)$$

In these coordinates the energy H of the system is given by

$$H=\sum\tfrac{1}{4}Lca_s^2+\sum\tfrac{1}{4}Llb_s^2. \quad (4)$$

Application of the statistical mechanical considerations appropriate to a system in small interaction with its

[1] J. E. Rhodes, Jr., Phys. Rev. **93**, 1 (1954).
[2] C. W. McCombie, Repts. Progr. Phys. **16**, 266 (1953).
[3] A. Einstein, Ann. Physik **19**, 1371 (1906).
[4] S. Chandrasekhar, Revs. Modern Phys. **21**, 383 (1949).

surroundings then yields the equipartition results

$$\tfrac{1}{4}Lc\langle a_s^2\rangle_{Av}=\tfrac{1}{2}kT, \quad \tfrac{1}{4}Ll\langle b_s^2\rangle_{Av}=\tfrac{1}{2}kT \tag{5}$$

if quantum effects, important only for very high frequency modes, are ignored.

The precise significance of these equations must be carefully considered. At any instant one can, at least in principle, carry out the indicated Fourier analysis and find the values of the coordinates. The electrical fluctuation energy is then given by (4), and the sum of the terms in a_s and b_s is taken to be the energy in the sth mode at the instant considered. The Eqs. (5) assert that the average of this last energy over a large number of instants will be kT.

In all this there is no suggestion that a_s and b_s will vary sinusoidally with time as they would on an undamped line. That the variation is not of this nature will become clear when the power spectra of the fluctuations of a_s and b_s are obtained in the next section.

III. RANDOM FORCE METHOD

The application of the Nyquist random force method to the line with shorted ends will be made sufficiently clear by a calculation of the power spectrum of the fluctuations of a_s. To determine the contribution to this power spectrum from the fluctuating voltage associated with an element of resistance or leakage conductance, a voltage $A\exp(2\pi i\nu t)$ is supposed connected in series with the element and the resulting steady state variation in the coordinate a_s expressed in the form $Y(\nu)A\exp(2\pi i\nu t)$. The required contribution is then $4RkT|Y(\nu)|^2$, where R is the resistance of the element.

The following preliminary consideration enables one to determine easily the function $Y(\nu)$ corresponding to any element. Suppose that the line is undisturbed except that, by some means or other, discontinuities in current and voltage $\Delta I\exp(2\pi i\nu t)$ and $\Delta V\exp(2\pi i\nu t)$ are maintained at a point P a distance x from one end. If the resulting voltage on the line is $V(y)\exp(2\pi i\nu t)$, the resulting value of a_s will be given by

$$a_s=\frac{2}{L}\int_0^L V(y)\exp(2\pi i\nu t)\sin(\pi sy/L)dy. \tag{6}$$

Now, on either side of P the voltage will have the form given by (1), the constants A and B having different values on the two sides. Substituting an expression of the form implied by this for $V(y)$ in (6) and using the fact that

$$V(0)=V(L)=0, \tag{7}$$

one easily finds that a_s can be expressed in terms of ΔI and ΔV in the form

$$a_s=\frac{2}{L}\left|\gamma^2+\frac{\pi^2s^2}{L^2}\right|^{-1}\left\{\gamma Z\Delta I\sin\frac{\pi sx}{L}+\frac{\pi s}{L}\Delta V\cos\frac{\pi sx}{L}\right\}\exp(2\pi i\nu t). \tag{8}$$

This is our preliminary result.

Consider now the contribution to the power spectrum of a_s from the fluctuating voltage associated with an element of resistance $r\delta x$ at P. When a voltage $A\exp(2\pi i\nu t)$ is connected in series with this element, one has

$$\Delta V=A, \quad \Delta I=0.$$

Inserting these values in (8) and comparing the result with $Y(\nu)A\exp(2\pi i\nu t)$ yields the value of $Y(\nu)$ which has to be substituted in $4r\delta xkT|Y(\nu)|^2$ to give the required contribution. Integrating over x leads to the contribution to the power spectrum of a_s from all such elements

$$G_{a_s}^{(r)}=\frac{8kTr}{L}\left(\frac{\pi s}{L}\right)^2\left|\gamma^2+\frac{\pi^2s^2}{L^2}\right|^{-2}. \tag{9}$$

Such integrations are valid because all the elements of resistance will contribute independently to the fluctuations of the coordinate.

When a voltage $A\exp(2\pi i\nu t)$ is connected in series with the element of leakage resistance $(1/\rho\delta x)$ the current through the element will, when δx is so small that the impedance of the line is negligible compared with $(1/\rho\delta x)$, be given by $\rho\delta xA\exp(2\pi i\nu t)$. Thus one has to substitute in (8)

$$\Delta V=0, \quad \Delta I=A\rho\delta x.$$

The resulting expression for $Y(\nu)$ is then substituted in $4(1/\rho\delta x)kT|Y(\nu)|^2$ and the result integrated over x to give for the total contribution to the power spectrum of a_s from elements of leakage conductance

$$G_{a_s}^{(\rho)}=\frac{8kT\rho}{L}|Z\gamma|^2\left|\gamma^2+\frac{\pi^2s^2}{L^2}\right|^{-2}. \tag{10}$$

Adding together the power spectra (9) and (10) and using (2) to express the result explicitly in terms of ν gives

$$G_{a_s}(\nu)=\frac{8kT}{L}\times\frac{4\pi^2\nu^2l^2\rho+r(r\rho+\pi^2s^2/L^2)}{(4\pi^2\nu^2lc-r\rho-\pi^2s^2/L^2)^2+4\pi^2\nu^2(rc+\rho l)^2}. \tag{11}$$

When the damping is very small (i.e., both r and ρ very small) this power spectrum is appreciable only for frequencies near the undamped frequency of the mode. In general, however, it will be appreciable over a considerable range of frequencies. This bears out the assertion made at the end of Sec. II.

Integrating the power spectrum (11) over all frequencies gives for $\langle a_s^2\rangle_{Av}$:

$$\langle a_s^2\rangle_{Av}=\int_0^\infty G_{a_s}(\nu)d\nu=2kT/Lc. \tag{12}$$

The integral is evaluated by partial functions or contour integration. The result, which is independent of r and ρ, agrees with (5).

The power spectrum of b_s can be obtained similarly. In fact, one gets for this just the right-hand side of (11) with r and ρ interchanged and l and c interchanged. Consequently

$$\langle b_s^2\rangle_{Av}=2kT/Ll \tag{13}$$

again in agreement with (5).

For a line with open ends the coordinates a_s and b_s will be defined by

$$V=\sum a_s\cos(\pi sy/L),\quad I=\sum b_s\sin(\pi sy/L). \tag{14}$$

One finds that the power spectra and consequently the mean square values of a_s and b_s are again given by the expressions obtained above.

IV. LONG TRANSMISSION LINE

A very long transmission line with arbitrary terminal conditions can be treated as follows. The actual normal coordinates for the undamped system will depend on the capacities and inductances of the terminal impedances. Suppose that, in spite of this, one introduces coordinates defined by (3). Then, although the Eq. (7) is no longer imposed by the boundary conditions, the relation (8) will hold for all points P such that a disturbance of any frequency originating at P does not extend to the end of the line. Thus for a sufficiently long line, one may suppose (8) to hold at all points, the contribution of the end portions to the final integrals being a negligible fraction of the whole. The results (11), (12), and (13) will therefore still hold.

The same conclusion would follow if coordinates defined by (14) were introduced. Consequently, for a long line terminated in an arbitrary way, the equipartition results hold in their simple form.

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Statistical Mechanics of Mixtures of Isotopes*

G. V. CHESTER
Sterling Chemistry Laboratory, Yale University, New Haven, Connecticut
(Received June 7, 1955)

The excess free energy of a mixture of isotopes is expanded in a Taylor series in powers of the relative mass differences $(m_0-m_j)/m_j$, where m_j and m_0 are the masses of the particles of the jth component and of a reference isotope, respectively. This expansion is only useful if it assumed that all the particles in the mixture obey classical or Boltzmann statistics. When this assumption is made it is found that the linear term in the expansion vanishes identically. The second-order term has the form $\Sigma_i{}^0\Sigma_j{}^0x_ix_j(\lambda_i-\lambda_j)^2\cdot Q$, where Q is a universal function of m_0, T, and V, λ_j equals $(m_0-m_j)/m_j$, and x_j is the mole fraction of the jth component. This expression gives the explicit dependence of the excess free energy on the mole fractions and on the relative mass differences λ_j. From this result it can be shown, among other things, that a phase separation of the isotopes in a mixture should take place at a sufficiently low temperature. It can also be shown that there is an approximate law of corresponding states between different mixtures of isotopes. The theory is directly applicable to all solid mixtures and to fluid mixtures of the hydrogen isotopes. Unfortunately owing to the lack of experimental data it is impossible to test the theory rigorously. Finally it is shown how the theory can be used to interpret the behavior of He^3-He^4 solutions. When sufficient experimental data becomes available the theory should throw considerable light on the influence of quantum statistics on the properties of these solutions.

I. INTRODUCTION

THE purpose of this paper is to develop an approximate theory of mixtures of isotopes. It is of course well known[1] that if the atoms or molecules of a mixture of isotopes are assumed to obey the laws of classical mechanics then the thermodynamic properties of the mixture can easily be related to those of the pure isotopes. Such mixtures are in fact "ideal solutions." However, if we assume that the particles of the system obey the laws of quantum mechanics then it is impossible to express, in any general way, the properties of a mixture of isotopes in terms of those of its components. Now we know, that, except at very low temperatures, the physical properties of most substances can be accurately accounted for by means of classical statistical mechanics. Consequently the properties of solutions of isotopes will remain ideal until these temperatures are reached. If we limit our attention to the class of simple substances to which De Boer's quantum mechanical law of corresponding states[2] applies, then we can easily estimate the highest temperature at which any particular mixture can be expected to show deviations from ideal behavior. For example, we find that (gaseous) mixtures of helium and hydrogen isotopes are unlikely to show nonideal behavior above 30°K and 60°K, respectively. These two substances are, however, the only ones that show appreciable departures from classical behavior in the gaseous and liquid states; all

* Contribution number 1301 from the Sterling Chemistry Laboratory.

[1] E. A. Guggenheim, *Mixtures* (Cambridge University Press, Cambridge, 1952), Chap. 2.

[2] J. De Boer, Physica **14**, 139 (1948).

other substances solidify before any departures from classical behavior appear. We can therefore conclude that, apart from these substances, mixtures of isotopes will be ideal in the liquid and gaseous states.

The theory we shall develop is based on the following three assumptions.

(a) We shall assume that the atoms or molecules of the system (for simplicity we shall henceforth refer to these as "particles") are "rigid." That is to say we assume that the motions of the electrons and nuclei, relative to the mass center of one of the particles, are not influenced by the presence of the other particles in the system. If we make this assumption then the partition function of the system immediately factorizes into two factors, Z_1 and Z_2. Here Z_1 is the partition function for the motions of the electrons and nuclei relative to the mass centers of the particles and Z_2 is the partition function for the motions of the mass centers themselves. Now the variation in Z_1 due to the difference between the masses of the isotopes makes no contribution to the free energy of mixing.[1] We can therefore ignore Z_1 and concentrate our attention entirely on Z_2, which we shall write simply as Z.

(b) Next we shall assume that the particles of all the isotopes in the mixture obey classical or Boltzmann statistics. While this assumption is not generally justified there are two important cases in which it is likely that it can be regarded as an accurate approximation. First, we should expect it to be a good approximation for the hydrogen isotopes. This is because the different types of statistics obeyed by the molecules of these isotopes arise from their different spin states. Consequently if the different spin states of the molecules exert a marked influence on the properties of the isotopes then we should expect a considerable difference in the properties of the *ortho* and *para* species of any particular isotope. The properties of the two species are, however, very nearly identical; they amount at the most to about 1%. We can therefore assume that the properties of the hydrogen isotopes are very nearly independent of the spin states of their molecules; consequently we can assume that they obey Boltzmann statistics. The accuracy of this assumption can very easily be tested by comparing the properties of mixtures of hydrogen isotopes that have different *ortho-para* concentrations. If the properties of the mixtures depend strongly on the *ortho-para* concentrations of the various concentrations of the various components then we must conclude that our assumption is unjustified. Secondly we should expect the assumption to be an accurate approximation for mixtures of isotopes in the solid state. For, in the solid state the particles in the system can usually be regarded as being localized on lattice sites, and it is well known[2] that in such a system both Fermi-Dirac and Bose-Einstein statistics reduce to Boltzmann statistics.

[1] R. H. Fowler, *Statistical Mechanics* (Cambridge University Press, Cambridge, 1936), Chap. 2.

On the other hand, this assumption is unlikely to be an accurate approximation for liquid or gaseous mixtures of He^3 and He^4. This is because the properties of these substances appear to depend critically on the type of statistics obeyed by their atoms. However, we shall see in Sec. 4 that if we use the theory in an indirect manner we can throw considerable light on the behavior of these solutions.

(c) Finally we assume that the interaction potential between the particles of any isotope is independent of the mass of the isotope. This assumption, while very accurate, is not completely correct for the following reason. The interaction potential depends on the electron density distribution round the nuclei. Now the reduced electron masses enter into this distribution and these in turn depend on the mass of the nuclei. However if the mass of the nuclei is changed by an amount δM then the reduced electron masses will change by an amount of the order of $10^{-5}\,\delta M$. This change is so small that it can be neglected entirely and consequently the interaction potential between the particles can be considered to be independent of the mass of the nuclei.

In the next section we give the formal mathematical development of the theory. This is based on the perturbation expansion of the partition function Z.[4,5] The expansion parameters are the relative mass differences $(m_0-m_j)/m_j$, where m_j is the mass of the jth component and m_0 is the mass of the particles of a reference isotope. The way in which these parameters enter into the partition function can readily be seen from the Hamiltonian of the system. Suppose we have a mixture of s isotopes. Let m_i and $x_i=N_i/N$ be the mass and the mole fraction of the ith component. Then we can write the total Hamiltonian, H, for the mixture in the form,

$$H=-\sum_{i=1}^{s}(1/m_i)\sum_{j=1}^{N_i}(h^2/2)\nabla_{r_j}^2+V(\mathbf{r}_1\cdots\mathbf{r}_N), \quad (1.1)$$

where $\mathbf{r}_1\cdots\mathbf{r}_N$ are the coordinates of the N particles and $V(\mathbf{r}_1\cdots\mathbf{r}_N)$ is the total interaction potential of all the particles. From assumption (c) this potential is independent of the masses m_j. To simplify the notation we write $k_j=(-h^2/2)\nabla_{r_j}^2$ and set

$$K(N_i,m_i)=(1/m_i)\sum_{j}^{N_i}k_j.$$

With this notation the Hamiltonian becomes

$$H=\sum_{i=1}^{s}K(N_i,m_i)+V(\mathbf{r}_1\cdots\mathbf{r}_N). \quad (1.2)$$

[4] M. R. Schafroth, Helv. Phys. Acta 24, 645 (1951).

[5] G. V. Chester, Phys. Rev. 93, 606 (1954). The author wishes to thank Dr. Schafroth for pointing out to him that many of the results obtained in this paper had been found earlier by Dr. Schafroth in the paper quoted above.

Now,

$$K(N_i,m_i)=\frac{1}{m_i}\sum_{j=1}^{N_i}k_j$$

$$=\frac{1}{m_0}\sum_{j=1}^{N_i}k_j+\frac{m_0-m_i}{m_i}\cdot\frac{1}{m_0}\sum_{j=1}^{N_i}k_j, \quad (1.3)$$

and if we set $\lambda_i=(m_0-m_i)/m_i$ we have,

$$K(N_i,m_i)=K(N_i,m_0)+\lambda_iK(N_i,m_0).$$

The Hamiltonian can therefore be written in the form,

$$H=\sum_{i=1}^{s}K(N_i,m_0)+V+\sum_{i=1}^{s}\lambda_iK(N_i,m_0), \quad (1.4)$$

that is,

$$H=K(N,m_0)+V+\sum_{i=1}^{s}\lambda_iK(N_i,m_0). \quad (1.5)$$

Finally, if we define H_0 by the equation

$$H_0=K(N,m_0)+V, \quad (1.6)$$

then H becomes

$$H=H_0+\sum_{i=1}^{s}\lambda_iK(N_i,m_0). \quad (1.7)$$

The free energy to the system can therefore be considered as a function of the λ_i's and we can formally expand it as a Taylor series in these parameters. Clearly we may expect such a series to be physically useful as long as the λ_i are not too large. We can also express the Hamiltonian for N particles of one of the pure isotopes in a similar way. Thus,

$$H_i=H(N,m_i)=H_0+\lambda_iK(N,m_0), \quad (1.8)$$

and we can again expand the free energy in powers of λ_i. Our method is therefore as follows. We expand the free energy of the mixture in powers of the λ_i's and we also expand the free energy of each of the pure isotopes in a similar series. From these series we can at once obtain a formal expansion of the free energy of mixing in powers of the λ_i's. The first term in this expansion gives the ideal free energy of mixing. The term that is linear in the λ_i vanishes identically and we have to include terms that are quadratic in the λ_i in order to obtain any departures from ideal behavior. This term is calculated explicitly as a function of the x_i and λ_i. An important feature of this method is that no assumption is made about the state of aggregation of the mixture and the theory should therefore be equally applicable to solid, liquid and gaseous mixtures. In Sec. 3, we discuss the physical consequences of the expression we have obtained for the excess free energy of mixing. The principal results that emerge from this discussion are: (a) that mixtures of isotopes should exhibit very nearly ideal behavior, this conclusion follows directly from the fact that the excess free energy depends only on second and higher powers of the relative mass differences, (b) that the free energy of mixing has the same dependence on the mole fractions as that of a "regular" solution,[6] (c) that at sufficiently low temperatures the excess free energy of mixing is always positive, consequently we expect a phase separation of the isotopes to take place at some temperature above absolute zero, and (d) that there exists an approximate law of corresponding states between mixtures of different isotopes. Unfortunately, we have not been able to express the excess free energy in terms of the properties of pure isotopes. The best we have been able to do is to obtain an upper bound for the excess free energy in terms of measurable quantities. There appears to be very little experimental data with which to compare the theory. In fact, only a qualitative comparison can be made; this comparison is however satisfactory. We conclude this section by discussing the experimental data that are required to test the theory rigorously.

In Sec. 4, we turn to a discussion of He^3-He^4 solutions. Our aim here is to throw some light on the influence of quantum statistics on the properties of these solutions. Now the theory we have developed is only valid if the particles of all the isotopes in the mixture obey Boltzmann statistics. Therefore if we compare the predictions of our theory with the experimental properties of He^3-He^4 solutions we should be able to show how the properties of these solutions are influenced by the different types of quantum statistics obeyed by He^3 and He^4 atoms. Thus if the experimental properties are in good agreement with the predictions of the theory then we can conclude that the properties of the solutions are not appreciably influenced by the different types of statistics obeyed by the atoms. On the other hand if there are large discrepancies between the predictions and the experimental data then we must draw the opposite conclusion. What little data are available indicates that at about 3°K the predictions and experimental data are in qualitative agreement but that at lower temperatures there are large discrepancies.

The only other work on the theory of isotopes appears to be that due to Prigogine.[7-9] Although his method of approach is very different from ours he has obtained very similar physical results. In Sec. 5 we shall compare these with our own.

2. FREE ENERGY OF MIXING

We consider a mixture of s components each of which is an isotope of the same substance. If m_i and x_i $(=N_i/N)$ are the mass and the mole fraction of the ith component then we have seen that the Hamiltonian of this system can be written in the form,

$$H=H_0(N,m_0)+\sum_{i=1}^{s}\lambda_iK_i, \quad (2.1)$$

[6] E. A. Guggenheim, reference 1, Chap. 4.
[7] Prigogine, Bingen, and Jeener, Physica **20**, 383 (1954).
[8] I. Prigogine and J. Jeener, Physica **20**, 516 (1954).
[9] Prigogine, Bingen, and Bellemans, Physica **20**, 633 (1954).

where for convenience we have set $K_i=K(N_i,m_0)$. Likewise the Hamiltonian of the ith component can be written,

$$H_i=H_0(N,m_0)+\lambda_i K. \tag{2.2}$$

Now the free energy of the mixture, which we denote by $F_m(N_1\cdots N_s,m_0,\lambda_1\cdots\lambda_s)$ is given by

$$F_m(N_1\cdots N_s,m_0,\lambda_1\cdots\lambda_s)=-kT\log Z, \tag{2.3}$$

where

$$Z=\mathrm{Tr}[\exp(-\beta H)]. \tag{2.4}$$

In this last equation $\beta=1/kT$, T is the absolute temperature, k is Boltzmann's constant, and the symbol $\mathrm{Tr}(A)$ stands for the trace of the operator A. Since we have assumed that all the different types of particles in the mixture obey Boltzmann statistics all the traces that occur in the theory must be formed with unsymmetrized wave functions. We now formally expand F_m in a Taylor series in the parameters λ_i. Thus,

$$F_m=F_m(N_1\cdots N_s,m_0,0)+\sum_{i=1}^{s}\lambda_i[\partial F_m/\partial\lambda_i]_{\lambda=0}$$

$$+\tfrac{1}{2}\sum_{i=1}^{s}\sum_{j=1}^{s}\lambda_i\lambda_j[\partial^2F_m/\partial\lambda_i\partial\lambda_j]_{\lambda=0}+O(\lambda^3), \tag{2.5}$$

where λ stands for the set $\lambda_1\cdots\lambda_s$. We now calculate the derivatives of F_m in terms of those of Z. Using Eq. (2.3) we find that

$$[\partial F_m/\partial\lambda_i]_{\lambda=0}=-[kT/Z(m_0,0)][\partial Z/\partial\lambda_i]_{\lambda=0}, \tag{2.6}$$

and

$$[\partial^2F_m/\partial\lambda_i\partial\lambda_j]_{\lambda=0}=(kT/Z^2)[\partial Z/\partial\lambda_i\cdot\partial Z/\partial\lambda_j]_{\lambda=0}$$
$$-(kT/Z)[\partial^2Z/\partial\lambda_i\partial\lambda_j]_{\lambda=0}. \tag{2.7}$$

We next calculate the derivatives of Z with respect to the λ_i by means of the perturbation expansions for the partition function[5]; we find that,

$$[\partial Z/\partial\lambda_j]_{\lambda=0}=-\beta\sum_l(K_j)_{ll}\exp(-\beta E_l), \tag{2.8}$$

and

$$[\partial^2Z/\partial\lambda_i\partial\lambda_j]_{\lambda=0}=\beta^2\sum_{l_1}\sum_{l_2}(K_i)_{l_1l_2}(K_j)_{l_2l_1}G_{l_1l_2}, \tag{2.9}$$

where $G_{l_1l_2}$ is given by,

$$G_{l_1l_2}=[\exp(-\beta E_{l_1})-\exp(-\beta E_{l_2})]/\beta(E_{l_2}-E_{l_1}). \tag{2.10}$$

In these equations, the E_{l_i} are the energy eigenvalues of the reference system of N particles of mass m_0 and the matrix elements $(K_i)_{l_1l_2}$ are formed with the corresponding wave functions. It is now convenient to define the following two averaging operations,

$$\langle A\rangle=\sum_l(A)_{ll}\exp(-\beta E_l)/\sum_l\exp(-\beta E_l), \tag{2.11}$$

and

$$\langle A,B\rangle=\sum_{l_1}\sum_{l_2}(A)_{l_1l_2}(B)_{l_2l_1}G_{l_1l_2}/\sum_l\exp(-\beta E_l). \tag{2.12}$$

The matrix elements $(K_j)_{l_1l_2}$ in Eqs. (2.8) and (2.9) are given by equations of the form,

$$(K_j)_{l_1l_2}=\sum_{t=1}^{Nj}(k_t)_{l_1l_2}, \tag{2.13}$$

where we have absorbed the factor $(m_0)^{-1}$ into k_t. If we combine Eqs. (2.11), (2.12), and (2.13) with Eqs. (2.8) and (2.9) we find that

$$\left[\frac{1}{Z}\frac{\partial Z}{\partial\lambda_i}\right]_{\lambda=0}=-\beta\langle K_i\rangle=-\beta\sum_{t=1}^{N_i}\langle k_t\rangle, \tag{2.14}$$

and

$$\left[\frac{1}{Z}\frac{\partial^2Z}{\partial\lambda_i\partial\lambda_j}\right]_{\lambda=0}=\beta^2\langle K_i,K_j\rangle=\beta^2\sum_{r=1}^{N_i}\sum_{t=1}^{N_j}\langle k_r,k_t\rangle. \tag{2.15}$$

Now $\langle k_i\rangle$ is the average value of the kinetic energy of the ith particle, averaged over all accessible states of the reference system. It is therefore independent of the suffix i; to indicate this we set $\langle k_i\rangle=\langle k_\alpha\rangle$, where α refers to any particle of the reference system. Similarly $\langle k_r,k_t\rangle$; $r\neq t$ is independent of the pair of particles that we chose and we can set it equal to $\langle k_\alpha,k_\beta\rangle$; $\alpha\neq\beta$. The suffixes r and t can, however, be equal and in this case we have to set $\langle k_r,k_r\rangle$ equal to $\langle k_\alpha,k_\alpha\rangle$. Using this notation we can write Eqs. (2.14) and (2.15) in the following way,

$$(1/Z\partial Z/\partial\lambda_i)_{\lambda=0}=-\beta N_i\langle k_\alpha\rangle, \tag{2.16}$$

and

$$(1/Z\partial Z/\partial\lambda_i\partial\lambda_j)_{\lambda=0}=\beta^2[N_iN_j\langle k_\alpha,k_\beta\rangle+\delta_{ij}N_i\Delta], \tag{2.17}$$

where Δ is given by,

$$\Delta=\langle k_\alpha,k_\alpha\rangle-\langle k_\alpha,k_\beta\rangle. \tag{2.18}$$

Substituting Eqs. (2.16) and (2.17) into Eqs. (2.6) and (2.7) we get,

$$(\partial F_m/\partial\lambda_i)_{\lambda=0}=N_i\langle k_\alpha\rangle=x_i\langle K\rangle, \tag{2.19}$$

and

$$(\partial^2F_m/\partial\lambda_i\partial\lambda_j)_{\lambda=0}$$
$$=\beta[N_iN_j\langle k_\alpha\rangle^2-N_iN_j\langle k_\alpha,k_\beta\rangle-\delta_{ij}N_i\Delta]$$
$$=\beta[x_ix_j\langle K\rangle^2-x_ix_j(\langle K,K\rangle-N\Delta)-\delta_{ij}x_iN\Delta]. \tag{2.20}$$

These last two equations lead to the following expansion for the free energy of the mixture

$$F_m(N_1\cdots N_s,m_0,\lambda_1\cdots\lambda_s)$$
$$=F_m(N_1\cdots N_s,m_0,0)+(\sum_{i=1}^{s}x_i\lambda_i)\langle K\rangle$$
$$+(\beta/2)(\sum_{i=1}^{s}\sum_{j=1}^{s}\lambda_i\lambda_jx_ix_j)[\langle K\rangle^2-(\langle K,K\rangle-N\Delta)]$$
$$-(\beta/2)(\sum_{i=1}^{s}\lambda_i^2x_i)N\Delta+O(\lambda^3). \tag{2.21}$$

Using exactly the same methods we can expand the free energy, F_i, of one mole of a pure isotope of mass m_i, thus,

$$F_i(N,m_0,\lambda_i)=F_i(N,m_0,0)+\lambda_i\langle K\rangle+(\beta/2)\lambda_i^2 \times[\langle K\rangle^2-(\langle K,K\rangle-N\Delta)]-(\beta/2)\lambda_i^2N\Delta. \quad (2.22)$$

We now have to examine the function

$$F_m(N_1\cdots N_s,m_0,0).$$

It is given by the equation,

$$F_m(N_1\cdots N_s,m_0,0)=-kT\log Z(N_1\cdots N_s,m_0,0),$$

where

$$Z(N_1\cdots N_s,m_0,0)=[\prod_j^s(N_j)!]^{-1}\sum_l\exp(-\beta E_l).$$

The factor

$$[\prod_{j=1}^s(N_j)!]^{-1}$$

in the last equation arises because the particles of the different isotopes in the mixture were originally distinguishable by their different masses. We can therefore write

$$Z(N_1\cdots N_s,m_0,0) =[N!/\prod_j(N_j)!](N!)^{-1}\sum_l\exp(-\beta E_l),$$

and consequently,

$$F_m(N_1\cdots N_s,m_0,0) =F(N,m_0,0)+NkT\sum_{i=1}^s x_i\log x_i. \quad (2.23)$$

The excess free energy of mixing ΔF^E, is defined by the equation,

$$\Delta F^E=F_m-\sum_{i=1}^s x_iF_i-NkT\sum_{i=1}^s x_i\log x_i. \quad (2.24)$$

Therefore from Eqs. (2.21), (2.22), and (2.23) we find that,

$$\Delta F^E=(\beta/2)(\sum_{i=1}^s\sum_{j=1}^s x_ix_j\lambda_i\lambda_j-\sum_{i=1}^s x_i\lambda_i^2) \times[\langle K\rangle^2-(\langle K,K\rangle-N\Delta)]+O(\lambda^3). \quad (2.25)$$

This equation gives the explicit dependence of ΔF^E on the x_i and λ_i. It is clear from the method of derivation that it would be a simple, though tedious, matter to calculate as many terms in the expansion as we wish.

The first factor on the right of Eq. (2.25) is easily shown to be equal to

$$-\sum_{i<j}^s x_ix_j(\lambda_i-\lambda_j)^2,$$

and is therefore negative for all values of the x_i and λ_j. We now shorten the notation by writing

$$\Delta F^E=c(x,\lambda)Q(T,V,m_0), \quad (2.26)$$

where x and λ stand for the sets (x_i) and (λ_i) respectively, and $c(x,\lambda)$, $Q(T,V,m_0)$ are given by

$$c(x,\lambda)=\sum_{i<j}^s x_ix_j(\lambda_i-\lambda_j)^2, \quad (2.27)$$

and

$$Q(T,V,m_0)=(2kT)^{-1}[\langle K,K\rangle-\langle K\rangle^2-N\Delta]. \quad (2.28)$$

The function $Q(T,V,m_0)$ is a universal function of T,V and m_0; that is, it does not depend on either the x_i or λ_i. Unfortunately, we have not been able to express it in terms of the properties of the pure isotopes. However the following equation can easily be derived:

$$\langle K,K\rangle-\langle K\rangle^2=-kT[\partial^2F_i/\partial\lambda_i^2]_{\lambda_i=0}. \quad (2.29)$$

Now, since we can also show that $\langle K,K\rangle-\langle K\rangle^2$ and Δ are always positive, we find that ΔF^E must obey the following inequality:

$$\Delta F^E\leq-\tfrac{1}{2}[\partial^2F_i/\partial\lambda_i^2]_{\lambda_i=0}. \quad (2.30)$$

This equation provides us with an upper bound for ΔF^E; in the next section we shall show how we can determine $[\partial^2F_i/\partial\lambda_i^2]_{\lambda_i=0}$, approximately, from experiment.

Finally, it is easily shown that the "fluctuation" term Δ tends to zero as $T\to0$. The sum of the other two terms in Q approaches $\sum_{l>0}|K_{l0}|^2/(E_l-E_0)$, as $T\to0$, and the excess free energy is therefore positive at absolute zero. This completes the formal part of our work.

3. PHYSICAL CONSEQUENCES OF THE THEORY

In this section we shall first discuss the physical consequences of the formulas we have derived and afterwards we shall compare the predictions of the theory with experiment.

The fundamental formula for ΔF^E is

$$\Delta F^E=c(x,\lambda)Q(T,V,m_0), \quad (3.1a)$$

where

$$c(x,\lambda)=\sum_{i<j}^s x_ix_j(\lambda_i-\lambda_j)^2, \quad (3.1b)$$

and

$$Q(T,V,m_0)=\tfrac{1}{2}\beta[\langle K,K\rangle-\langle K\rangle^2-N\Delta]. \quad (3.1c)$$

From these equations we can draw the following conclusions.

(a) The foregoing formulas were derived without making any assumptions about the state of aggregation of the system; they therefore apply quite generally to gaseous, liquid and solid mixtures.

(b) Equation (2.27) shows that ΔF^E depends only on second and higher order terms in the mass differences

λ_i. This makes it seem likely that mixtures of isotopes that obey Boltzmann statistics will, in general, show only very small deviations from ideal behavior. We shall see that this conjecture is confirmed by what little experimental data is available.

(c) We have seen that the excess free energy of mixing is positive at absolute zero. On the other hand the ideal free energy of mixing goes to zero as $T\to 0$. Therefore at some temperature above absolute zero we should expect a phase separation of the isotopes in the mixture. The highest temperature, T_c, at which this separation can occur is found, by the usual methods,[9] to be given by the conditions,

$$RT_c=\tfrac{1}{2}(\lambda_1-\lambda_2)^2Q(T_c,V,m_0), \quad x_1=x_2=\tfrac{1}{2}, \quad (3.2)$$

where we have specialized to a binary mixture. Equation (3.2) must be regarded as a transcendental equation for T_c; clearly as long as Q is negative it has no solution. It is interesting to note that this phase separation is a purely quantum mechanical phenomenon. If the particles of the mixture were assumed to obey the laws of classical mechanics then the solution would always be ideal.

(d) Next we notice that Eq. (3.1) give the explicit dependence of ΔF^E on the mole fractions x_i. It is clear from this equation that if we know ΔF^E for any one set of concentrations x then we can at once predict it for any other set x'. For we have at once that

$$\Delta F^E(T,V,x')=[c(x',\lambda)/c(x,\lambda)]\Delta F^E(T,V,x). \quad (3.3)$$

(e) Equation (3.1) also gives the explicit dependence of ΔF^E on the relative mass differences λ_i. Therefore if we have determined ΔF^E experimentally for one solution with mass differences λ and mole fractions x we can at once predict ΔF^E for any other solution with mass differences λ' and mole fractions x'. We have in fact,

$$\Delta F^E(T,V,x',\lambda')=[c(x',\lambda')/c(x,\lambda)]\Delta F^E(T,V,x,\lambda). \quad (3.4)$$

This last equation really expresses a law of corresponding states for mixtures of isotopes. In fact Eq. (3.2) tells us that if we know ΔF^E for a solution and divide it by the appropriate value of $c(x,\lambda)$ then we obtain a universal function of T, V, and m_0. Clearly since these remarks apply to any other thermodynamic property of the solution we are justified in speaking of a law of corresponding states.

(f) We can easily show that $\Delta F^E\to 0$ as $T\to\infty$, or as $k\to 0$. Therefore we should expect ΔF^E to increase as the temperature of the mixture is lowered. However, it appears that the properties of pure isotopes become very nearly identical when we pass from the liquid to the solid phase. This is strikingly shown by the plots of the specific heats of the hydrogen isotopes that appear in the article by Wooley, Scott, and Brickwedde.[10] It seems likely therefore that ΔF^E will decrease sharply as the mixture solidifies.

(g) Equation (3.16) shows that ΔF^E has the same dependence on the mole fractions x_i as a "regular" solution. Our formula differs from that for a regular solution in that Q is function both of T and V (or P).

(h) Finally we note that Eq. (2.30) provides an upper bound for ΔF^E. To our degree of approximation we find that $[\partial^2 F_i/\partial\lambda_i^2]_{\lambda_i=0}$ is given by the equation,

$$\left[\frac{\partial^2 F_i}{\partial\lambda_i^2}\right]_{\lambda_i=0}=\frac{1}{\lambda_i\lambda_j(\lambda_i-\lambda_j)}[\lambda_i F_j-\lambda_j F_i-(\lambda_i-\lambda_j)F_0], \quad (3.5)$$

where F_j is the free energy of a pure isotope of mass m_j. In principle therefore we can determine the upper bound for ΔF^E provided we know the free energies of two pure isotopes together with that of the reference species.

The only published data on mixtures of isotopes other than helium is that due to Kerr[11] and Hoge and Arnold.[12] Kerr measured the excess molar volume of a mixture of T_2, DT, and D_2 and found that it was about 0.6% of the total volume of the mixture. Unfortunately he worked with only one solution at fixed composition and we therefore cannot test either the dependence on the mole fractions or the approximate law of corresponding states predicted by the theory. However, the fact that the excess molar volume is so small is in good agreement with our prediction that isotopic solutions should be very nearly ideal. Hoge and Arnold measured the dew points pressures for two H_2-D_2 mixtures and for three H_2-HD mixtures. From their data, we can calculate the excess free energy of mixing for the solutions. The results of these calculations are tabulated in Table I. First we note that the values of ΔG^E range from 2% to 4% of the ideal free energies of mixing and are therefore comparitively small—as we have predicted. Let us take HD as the reference isotope. Then we find that $\lambda_1=-\frac{1}{3}$ and $\lambda_2=\frac{1}{3}$, for H_2 and D_2, respectively. Now solutions 3 and 5 both refer to the same temperature and roughly the same pressure. Consequently, from Eq. (3.1), we find that

$$\Delta G^E(3)/\Delta G^E(5)=(x_1x_2)_3/(x_1x_2)_5\simeq 1.2,$$

which is in satisfactory agreement with experiment. Likewise solutions 1 and 3 refer to roughly the same temperature and pressure and we find that we should have,

$$\Delta G^E(3)/\Delta G^E(1)=\lambda_1^2(x_1x_2)_3/(\lambda_1-\lambda_2)^2(x_1x_2)_1\simeq 0.5.$$

Again the agreement with experiment is satisfactory. Unfortunately this is as far as we can go with this type of comparison.

It is fairly clear from our discussion that the two

[9] Wooley, Scott, and Brickwedde, J. Research Natl. Bur. Standards **41**, 379 (1948).

[11] E. C. Kerr, J. Am. Chem. Soc. **74**, 824 (1952).

[12] H. J. Hoge and R. D. Arnold, J. Research Natl. Bur. Standards **47**, 63 (1951).

TABLE I. Excess free energy of mixing for solutions investigated by Hoge and Arnold (reference 12).

Solution No.	1	2	3	4	5
Temperature (T) °K	19.00	20.00	18.00	17.03	18.00
Pressure (P) mm/Hg	234±2	322±2	280±1	158±1	286±1
Mole fraction H_2	0.499	0.460	0.503	0.503	0.752
Mole fraction HD	0.004	0.004	0.496	0.686	0.268
Mole fraction D_2	0.497	0.556	0.001	0.001	0.000
Excess free energy ΔG^E in cal/mole	1.65±0.8	1.15±0.5	0.80±0.4	0.70±0.3	0.80±0.4

most satisfactory methods of testing the predictions of the theory would be (a) to test the dependence of the excess molar quantities on the mole fractions of the components in the mixture and (b) to test the approximate law of corresponding states for mixtures of different isotopes. To test the dependence of the excess molar quantities on the mole fractions we require data, at fixed T and P (or V), and for various concentrations. On the other hand to test the law of corresponding states we require data on a number of solutions of different isotopes, again at fixed T and P (or V). Since we can form three binary solutions from H_2, D_2, and HD it is likely that a thorough study of solutions of these three isotopes will provide us with sufficient data with which to test the theory.

4. He^3-He^4 SOLUTIONS

There is now a variety of evidence which indicates that the properties of the two pure isotopes of helium, He^3 and He^4, are to a large extent dependent on the type of statistics obeyed by their atoms. For example the specific heats and hydrodynamical flow properties are very different, at least for all temperatures for which measurements are available. The theory that we developed in the last two sections was, however, based on the assumption that all the particles in the mixture under consideration obeyed Boltzmann statistics. We therefore cannot expect to apply the theory directly to solutions of He^3 and He^4 isotopes. However, just because of this fact we can say that any deviations of the properties of He^3-He^4 solutions from those predicted by the theory are likely to be due to the fact that He^3 and He^4 atoms obey quantum statistics rather than Boltzmann statistics. We must therefore consider what properties we should predict for He^3-He^4 solutions.

The two most important predictions of the theory are, first, that the excess free energy of mixing depends only on second and higher powers of the relative mass differences and second, that the excess free energy depends on the mole fractions through the factor x_1x_2. As we have already pointed out, the first prediction indicates that the excess molar quantities should be small, and this prediction was born out by the data we quoted on the hydrogen isotopes. Now it is easily seen that the excess free energy of a mixture of isotopes whose particles obey quantum statistics rather than classical statistics *depends on both zero and first order terms in the mass differences*. Consequently we should expect that the excess molar quantities will be much larger for this type of solution. However, as the temperature of the solution is raised the influence of the different types of quantum statistics will decrease and it is likely that at sufficiently high temperatures the excess molar quantities will become of the same order of magnitude as those found in mixtures of hydrogen isotopes. Unfortunately there appears to be very little data available on He^3-He^4 mixtures. The only data that extend to temperatures well above the λ temperature are those due to Weinstock, Osborne, and Abraham.[13] They found that the vapor pressure for a 20% mixture of He^3 in He^4 was 6%, 20%, and 80% greater than that for an ideal solution at temperatures of 3°K, 2°K, and 1.2°K respectively. These values are in good qualitative agreement with the ideas presented in the foregoing. They seem to indicate that at about 3°K the different types of quantum statistics have very little influence on the properties of He^3 and He^4 mixtures. A much more satisfactory method of determining the influence of quantum statistics on these solutions would be to test, at fixed temperature and pressure (or volume), the dependence of some excess molar quantity on the mole fractions. Unfortunately there is insufficient data to enable us to do this satisfactorily. However, Nanda[14] has analyzed the rather scanty data due to Sommers,[15] on He^3-He^4 solutions above the λ temperature. He assumed that at these temperatures the solutions obey the composition law we have derived in this paper. We should note that the theory we have developed definitely shows that the nonideality parameter $W[=\frac{1}{2}(\lambda_1-\lambda_2)^2Q]$ is a function of both T and V (or P). We consequently cannot expect to fit the available data with a temperature independent W. Unfortunately, Sommer's data does not extend to temperatures above 2.2°K[16] and it

[13] Weinstock, Osborne, and Abraham, Phys. Rev. **77**, 400 (1950).

[14] V. S. Nanda, Phys. Rev. **97**, 571 (1955).

[15] H. S. Sommers, Phys. Rev. **88**, 113 (1952).

[16] In a recent issue of Zhur. Eksptl. i Teort. Fiz. **26**, No. 6 (1954). B. Eselson has published a large amoung of data on solutions ranging in concentration from 0.4% to 8.08% and from temperatures of 1.3°K to 3.2°K. Unfortunately his data disagrees both with that due to Sommers and with that due to Weinstock, Osborne, and Abraham. In particular he finds large negative deviations from ideality above the λ temperature and these deviations show no signs of decreasing at 3.0°K. This behavior also contradicts our theoretical predictions.

is doubtful whether our theory is applicable at temperatures as near the λ point as this. Nevertheless Nanda, found that the data leads to a fairly consistent value of W of 1.8 cals/mole. He also states that this value for W will account for the experimental data of Daunt and Tseng[17] on a 4% solution of He^3 in He^4. To check Nanda's calculations we have calculated W from Weinstock's data on a 20% solution and we find that this leads to a value of 2.0 cals/mole, at 2.2°K. Considering the uncertainty in the experimental measurements, this value can be considered to be in good agreement with Nanda's value.

When a He^3-He^4 solution is solidified by the application of pressure it is almost certainly correct to assume that the atoms become localized on lattice sites. Consequently, when we enumerate the states of the system it makes no difference whether we apply Fermi-Dirac or Bose-Einstein statistics to the particles. For in a localized system both methods of enumeration lead to the same result as that obtained by the application of Boltzmann statistics.[3,18] Our theory is therefore directly applicable to solid solutions of He^3 and He^4. Perhaps the best way to test the theory in this instance would be to measure some simple property, such as the velocity of sound or the specific heat at constant pressure. The dependence of both these quantities on the mole fractions can easily be calculated from the formula we have derived for ΔF^E. The predicted dependence can then be compared with that found experimentally.

So far our remarks have been entirely qualitative in nature. We shall now show how, in principle at least, a quantitative comparison can be made between the properties of a helium mixture and the properties of a hydrogen mixture. Let us consider Eq. (3.2)

$$\Delta F^E/c(x,\lambda)=Q(m_0,T,V). \tag{4.1}$$

We know from De Boer's work on the law of corresponding states[3] that the partition function Z depends only on the reduced variables T^*, V^*, and the parameter Λ^*. Here $T^*=kT/\epsilon$, $V^*=V/N\sigma^3$ and $\Lambda^*=h/\sigma(m\epsilon)^{\frac{1}{2}}$, where ϵ, σ, m have the usual significance. If we use the same methods as De Boer then we can easily show that $Q^*=Q/N\epsilon$, is also a function of T^*, V^*, and Λ^* alone. Therefore if we set $(\Delta F^E)^*=\Delta F^E/N\epsilon$, we can write Eq. (4.1) in the form

$$(\Delta F^E)^*/c(x,\lambda)=Q^*(T^*,V^*,\Lambda_0^*). \tag{4.2}$$

We now consider two different mixtures; (i) a mixture of hydrogen isotopes and (ii) a mixture of He^3 and He^4. Let us choose reference masses m_{01} for the first and m_{02} for the second. Then if we make the assumption that the atoms and molecules of all the isotopes in the two mixtures obey Boltzmann statistics we can write down two equations similar to Eq. (4.1), namely,

$$(\Delta F^E)_1^*/c(x,\lambda)=Q^*(T^*,V^*,\Lambda_{01}^*), \tag{4.3}$$

and

$$(\Delta F^E)_2^*/c(x',\lambda')=Q^*(T^*,V^*,\Lambda_{02}^*), \tag{4.4}$$

where x, λ refer to the hydrogen mixture, x', λ' refer to the helium mixture, and Λ_{01}^* and Λ_{02}^* are the appropriate values of Λ^*. Clearly the two functions on the right hand sides of Eqs. (4.3) and (4.4) will become identical if we chose m_{01} and m_{02} so that $\Lambda_{01}^*=\Lambda_{02}^*$. This implies that m_{01} and m_{02} must satisfy the equation,

$$m_{01}/m_{02}=(\epsilon_2/\epsilon_1)(\sigma_2/\sigma_1)^2\simeq 0.2. \tag{4.5}$$

The numerical value was obtained by substituting the known[19] values of ϵ_1, ϵ_2, σ_1, and σ_2. When this condition is fulfilled we can write,

$$(\Delta F^E)_1^*/c(x,\lambda)=(\Delta F^E)_2^*/c(x',\lambda'), \tag{4.6}$$

where of course $(\Delta F^E)_1^*$ and $(\Delta F^E)_2^*$ are to be regarded as a function of the reduced variables T^* and V^*. It is obvious from this last equation that a similar equation must hold between the reduced values of any other property of the two mixtures. In fact this equation merely expresses a general, though approximate, law of corresponding states between the properties of mixtures of isotopes of different substances; it can be regarded as a generalization of the more restricted law that we discussed in the last section. The reason why we are able to set up a law of corresponding states between the properties of quantum mechanical systems is that we have considerable freedom in the choice of the reference masses and m_{01} and m_{02}. As long as we chose them so that Eq. (4.5) is satisfied the above law will hold.

We can now use Eq. (4.5) to predict the properties of He^3-He^4 solutions from those of a hydrogen mixture. Let us suppose that we have measured a reduced excess quantity, $(\Delta X)_1^*$, as a function of T^* and P^* for a hydrogen mixture. Then we have at once that $(\Delta X)_2^*$ for a He^3-He^4 solution is given by

$$(\Delta X)_2^*=[c(x',\lambda')/c(x,\lambda)](\Delta X)_1^*. \tag{4.7}$$

But this prediction is only possible if we assume that the atoms or molecules of all the substances obey Boltzmann statistics. We can therefore say that any departures of $(\Delta X)_2^*$ from that predicted by the above equation are almost certain to be due to the different types of quantum statics obeyed by the He^3 and He^4 atoms. Equation (4.7) and other similar equations therefore enable us to determine, in principal at least, the influence of quantum statistics on the properties of He^3-He^4 solutions.

We must now consider whether such a comparison can be carried through in practice. There are two points to be considered in this connection. Firstly, Eq. (4.1)

[17] J. G. Daunt and T. P. Tseng (unpublished).

[18] This fact suggests that there may be a marked decrease in the deviations from ideal behavior of He^3-He^4 solutions as they solidify. It is also likely that the dependence of ΔF^E on the mole fractions will change as the solution solidifies.

[19] J. De Boer, Repts. Progr. Phys. 12, 305 (1948).

will be an accurate approximation as long as m_{01} and m_{02} are chosen so that (a) Eq. (4.5) is satisfied and (b) so that the relative mass differences λ and λ' for the two mixtures are as small as possible. The largest value of m_{02} would therefore appear to be about 5 amu, with this value λ_1' and λ_2' are $\frac{3}{4}$ and $\frac{1}{4}$, respectively. If m_{02} is about 5 amu then m_{01} must be about 1 amu to satisfy Eq. (4.5). We then find that for a H_2-HD mixture, λ_1 and λ_2 are $-\frac{1}{3}$ and $-\frac{2}{3}$, respectively. While these values are rather large it is unlikely that they are so large as to render the theory inapplicable. Secondly, if we wish to make an exact comparison between the properties of the two mixtures then the reduced thermodyhamic variables must have the same values. For example, if we consider properties under conditions of constant pressure and temperature then we must have $T_1^*=T_2^*$ and $p_1^*=p_2^*$. Using the known values of ϵ_1, ϵ_2, σ_1, and σ_2 we find that this means that $T_1=3.7T_2$ and $p_1=2.5p_2$. It should be a simple matter to fulfill these two conditions.

5. COMPARISON WITH PRIGOGINE'S THEORY

In a series of interesting papers[7-9] (hereafter referred to as I, II, and III) Prigogine has developed an approximate theory of solutions of isotopes. Papers I and II are concerned with the thermodynamic properties of mixtures of isotopes at absolute zero. The assumptions are made that the particles of the mixture are localized on lattice points and that they are coupled by means of harmonic forces. It is, of course, impossible to solve even this simple problem completely, except in the one dimensional case.[20] Prigogine therefore treats the three-dimensional case by means of perturbation theory. He finds that the first-order terms in the mass differences do not contribute to the energy of mixing and that the second order term is always as positive. This latter result leads at once to the prediction of a phase separation of the isotopes at some temperature above absolute zero. Both of these results are in complete agreement with our own; namely (ii) and (iii) of Sec. III. However, on the basis of these assumptions the energy of mixing is found to be only about 10^{-1} cal/mole. That is to say it is about ten times smaller than that found in mixtures of hydrogen isotopes. As Prigogine points out, this large discrepancy is almost certainly due to the very large anharmonic forces that come into play because of the higher zero point energy of these substances.

In paper III, Prigogine introduces a new assumption. This is that if by means of suitable external pressures we make the molar volumes of the components equal to that of the mixture then the free energy of the mixture should be equal to that of the components, apart from the usual configurational entropy term. This means that before we mix the two components we must apply a suitable positive pressure to one a negative pressure to the other. With this assumption the free energy of the mixture, $f_m(T,v)$, is given by the equation,

$$f_m(T,v)=x_1f_1(T,v)+x_2f_2(T,v)-RT\sum_{i=1,2}x_i\log x_i, \quad (5.1)$$

where $f_i(T,v)$ is the free energy of the ith component at a temperature T and molar volume v. If we expand each of the f_i in Eq. (5.1) in a Taylor series in powers of $(v-v_i)$ and substitute these expansions into the above expression for f_m then we can easily relate the properties of the mixture to those of its components. The principal results that follow from this treatment are as follows:

(a) The excess free energy is proportional to $(v_1-v_2)^2$, when terms of order $(v-v_i)^3$ are neglected, and is always positive. Again we should expect a phase separation of the isotopes. This result is in agreement with our own conclusions.

(b) The numerical value for the excess free energy is in good agreement with the experimental values derived from the data due to Hoge and Arnold.[13]

(c) The excess free energy is not a symmetrical function of the mole fractions x_1 and x_2. This result therefore is different from our own and consequently can be used to compare the two theories experimentally.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor J. G. Kirkwood and my colleagues for the numerous conversations I have had with them throughout the course of this work. The author also wishes to thank the United States Government for the award of a F. O. A. Research Fellowship.

[20] F. J. Dyson, Phys. Rev. **92**, 1331 (1954).

λ Transition in Liquid Helium

G. V. CHESTER
*Sterling Chemistry Laboratory, Yale University, New Haven, Connecticut**

(Received March 24, 1955)

It is shown, by means of an expansion formula, that Feynman's partition function for liquid helium exhibits a second-order lambda transition. The validity of the expansion is discussed and it is shown that it is very likely to be valid for the case under consideration. A discussion is also given of some general properties of Feynman's partition function.

1. INTRODUCTION

IN a recent publication[1] Feynman has proposed a comparatively simple approach to the problem of the phase transition in liquid helium. This approach consisted of constructing an alternative, exact, mathematical form for the partition function for liquid helium, and then using a series of intuitive physical arguments to derive from it a much simpler approximate partition function. We have discussed the derivation of this approximate partition function elsewhere[2] and we do not propose to enter into any further discussion of this aspect of Feynman's work in this paper. Unfortunately it is probably impossible, at least at the moment, to evaluate even this comparatively simple partition function. In his original paper[1] Feynman showed, using some rather crude approximations, that his partition function could exhibit a third order transition under conditions of constant volume. It is a simple matter to show that it also exhibits a third-order transition under conditions of constant pressure. Recently several authors[2–5] have advanced opinions as to whether or not Feynman's form for the partition function can yield a second-order transition. This question is of some importance because if the partition function does exhibit such a transition then it will at least be in qualitative agreement with experiment.[6] Kikuchi has in fact shown, on the basis of a simple lattice model, that the partition function can exhibit a second order transition. He pointed out, however, that this result is not completely rigorous because of the unrealistic nature of the lattice model.

The purpose of this paper is to present what we believe is a fairly rigorous proof that Feynman's partition function does exhibit a second-order transition. Besides doing this we shall, however, discuss some general properties of Feynman's partition function.

Our proof of the existence of a second-order transition is based on an expansion formula for the partition function, Z. This function can be written as,[2]

$$Z=\frac{(\lambda')^{3N}}{N!}\int\cdots\int \exp[-\beta V'(\beta)] \times\rho_0(\mathbf{r}_1\cdots\mathbf{r}_N)d\mathbf{r}_1\cdots d\mathbf{r}_N, \quad (1.1)$$

where $(\lambda')^2=\hbar^2/2\pi m'kT$, $\beta=1/kT$, m' is the effective mass introduced by Feynman and $V'(\beta)$ is the quasi interaction potential of the helium atoms.[3] The function $\rho_0(\mathbf{r}_1\cdots\mathbf{r}_N)$ is the probability density, in configuration space, for an ideal Bose Einstein gas of particles of mass m'. Finally the integral $\int\cdots\int\cdots d\mathbf{r}_1\cdots d\mathbf{r}_N$ indicates an integration over all the coordinates $\mathbf{r}_1\cdots\mathbf{r}_N$ of the helium atoms. Now Feynman has pointed out[1] that the essential property of the function $\exp(-\beta V')$ is that it must vanish whenever two or more helium atoms approach sufficiently close to one another. This simply means that configurations in which two or more atoms are very close together must be given zero weight in the partition function. Bearing this fact in mind we can construct a good approximation to the function $\exp(-\beta V')$ by assuming that $V'(\beta)$ is the potential energy for a system of N perfectly elastic spheres, the radii of which are independent of the temperature. This assumption does not imply any great loss of generality, for it preserves the essential property of the function and in addition any weak attractive or repulsive terms that may be present in $V'(\beta)$ can easily be taken into account by means of perturbation theory.[7] With this assumption we can write

$$\exp[-\beta V'(\beta)]=\prod_{i<j}^{N}[1+f(r_{ij})], \quad (1.2)$$

where $f(r_{ij})=f_{ij}$ is given by

$$\begin{aligned} f_{ij}&=-1; \quad r_{ij}\leqslant 2r_0 \\ &=0; \quad r_{ij}>2r_0, \end{aligned} \quad (1.3)$$

* Contribution No. 1283 from the Sterling Chemistry Laboratory.

[1] R. P. Feynman, Phys. Rev. **91**, 1291 (1953).

[2] G. V. Chester, Phys. Rev. **93**, 1412 (1954).

[3] D. ter Haar, Phys. Rev. **95**, 895 (1954).

[4] T. Matsubara, Busseiron Kenyku **72**, 78 (1954).

[5] R. Kikuchi, Phys. Rev. **96**, 563 (1954).

[6] W. H. Keesom, *Helium* (Elsevier Publishing Company, Inc., Amsterdam, 1942).

[7] In the second reference we showed how Feynman's partition function could be obtained from the full partition function by means of the semiclassical expansion in powers of $\hbar^2$. However, if the interaction potential is singular over regions of configuration space, as is the case for perfectly elastic spheres, this method breaks down. This is simply because for this type of potential it is impossible to derive the usual expansion in powers of $\hbar^2$. A very similar difficulty arises in the development of the perturbation expansion for the partition function. This has been discussed by the author in Phys. Rev. **93**, 606 (1954).

and r_0 is the radius of the elastic spheres. The expansion of the partition function which we shall construct is essentially an expansion in powers of Nr_0^3/V, where N/V is the density of the system. It should be pointed out that this expansion of Z is not identical with virial expansion and is therefore not necessarily divergent for a liquid phase. We shall return to this point later. In Sec. 2 we shall show how this expansion can be constructed and we shall calculate the first two terms of it explicitly.

In Sec. 3 we study the thermodynamic properties of the first two terms in this expansion and show that they lead to a second-order transition. Since this result is based on the first two terms in the expansion for Z we must try to decide whether or not it is true for the exact form for Z. We therefore conclude this section with a discussion of (a) the validity of the expansion formula itself and (b) the validity of breaking off the expansion after only two terms. The essential result that emerges from this discussion is that the results we have derived by our approximate method are very likely to be true for the exact form for Z. We have not, however, been able to construct a completely rigorous argument to show this.

Finally in Sec. 4 we discuss some general properties of Feynman's partition function.

2. THE EXPANSION FORMULA

We first construct an explicit expression for $\rho_0(\mathbf{r}_1\cdots\mathbf{r}_N)$. This function is the unnormalized probability density, in configuration space, for a system of N noninteracting particles of mass m'. If the eigenvalues of Schrödinger's equation for such a system are E_l and if ϕ_l are the corresponding, correctly symmetrized, wave functions then

$$[(\lambda')^{3N}/N!]\rho_0(\mathbf{r}_1\cdots\mathbf{r}_N)=\sum_l \phi_l^*\phi_l \exp[-\beta E_l]. \quad (2.1)$$

Substituting this expression in Eq. (1.1), we get,

$$Z=\sum_l\left[\int\cdots\int\phi_l^* \exp(-\beta V')\phi_l d\mathbf{r}_1\cdots d\mathbf{r}_N\right] \times\exp(-\beta E_l), \quad (2.1)$$

and if we define R_{ll} by the equation

$$R_{ll}=\int\cdots\int\phi_l^* \exp[-\beta V']\phi_l d\mathbf{r}_1\cdots d\mathbf{r}_N, \quad (2.2)$$

then

$$Z=\sum_l R_{ll}\exp[-\beta E_l]. \quad (2.4)$$

This last equation can be rewritten in the form

$$Z=Z_0\langle R_{ll}\rangle^0, \quad (2.5)$$

where

$$Z_0=\sum_l \exp(-\beta E_l), \quad (2.6)$$

and

$$\langle R_{ll}\rangle^0=\sum_l R_{ll}\exp(-\beta E_l)/\sum_l \exp(-\beta E_l). \quad (2.7)$$

It follows at once that the free energy F can be written as

$$F=F_0+F_1, \quad (2.8)$$

where

$$F_0=-kT\log Z_0;\quad F_1=-kT\log\langle R_{ll}\rangle^0. \quad (2.9)$$

So far we have made no approximations. We now wish to develop an expansion formula for F_1. To do this we substitute into Eq. (2.3) the expression given by Eq. (1.2) for $\exp(-\beta V')$. Thus,

$$R_{ll}=\int\cdots\int\phi_l^*\phi_l\prod_{i<j}(1+f_{ij})d\mathbf{r}_1\cdots d\mathbf{r}_N. \quad (2.10)$$

The integral on the right hand side of Eq. (2.10) is very similar to the configuration integral that arises in the classical theory of compressed gases.[8] In fact if we put $\phi_l^*\phi_l=1$, then the integrals become identical. This suggests that we can probably obtain an expansion for F_1 by using the same techniques[8] that are used to develop the cluster expansion of the classical configuration integral. This surmise proves to be correct and it will clarify our work if we briefly recall the main steps in the calculation of the first two terms in the classical cluster expansion.

In the classical case we have to consider the integral

$$K=\int\cdots\int\prod_{i<j}(1+f_{ij})d\mathbf{r}_1\cdots d\mathbf{r}_N. \quad (2.11)$$

The first step is to expand the product $\prod_{i<j}(1+f_{ij})$ as a polynomial, the terms of which are sums of products of the f_{ij}. Thus,

$$\prod_{i<j}(1+f_{ij})=1+\sum_{i<j}f_{ij}+\sum_{i_1<j_1}\sum_{i_2<j_2}f_{i_1j_1}f_{i_2j_2}\cdots. \quad (2.12)$$

We next integrate this expansion over all the coordinates. The first term is trivial and leads at once to a term V^N. To obtain the first-order correction to this term we simply have to select from each term in the polynomial just those terms in which there are no f_{ij}'s with common suffices. For example in the first sum, $\sum_{i<j}f_{ij}$, as there are no terms with more than one f_{ij} we keep all the terms. In the second sum, however, we select terms like $f_{ij}f_{kl}$ but reject terms like $f_{ij}f_{jk}$. In the rth sum in the polynomial each term consists of r f_{ij}'s multiplied together, and as we only retain terms with no common suffices this means that we have $2r$ different coordinates appearing in the f_{ij}, and $N-2r$ that do not appear. Therefore when we integrate over all the coordinates, $N-2r$ of these integrations can be carried out at once to yield a factor V^{N-2r}. The remaining $2r$ integrations can also be easily performed—simply because there are no common suffixes in the product of the f_{ij}'s—and yield a factor z^r, where z^r is

[8] R. H. Fowler, *Statistical Mechanics* (Cambridge University Press, Cambridge, 1936), Chap. 8.

given by

$$z=\int_{\rho} f(\boldsymbol{\rho})d\boldsymbol{\rho}, \tag{3.13}$$

and $\boldsymbol{\rho}$ is equal to $\mathbf{r}_i-\mathbf{r}_j$. Consequently from any one term in the rth sum we get a contribution $V^{N-r}z^r$. To get the complete contribution to K from the sum we multiply this typical contribution by the number of terms we retain. This number is just the number of ways of picking r different pairs of suffixes from a total of N suffixes. If this number is denoted by $P(r,N)$, then the total contribution to K from the rth sum is $P(r,N)V^{N-r}z^r$. We now have to sum this expression over all values of r from 0 to $N/2$ to obtain, to our degree of approximation, the complete expression for K, namely,

$$K=\sum_{r=0}^{N/2} P(r,N)V^{N-r}z^r=V^N\sum_{r=0}^{N/2} P(r,N)N^{-r}2^r x^r, \tag{2.14}$$

where

$$x=\frac{N}{2V}\int_{\rho} f(\boldsymbol{\rho})d\boldsymbol{\rho}, \tag{2.15}$$

and depends on N and V only through the density N/V. It is now a fairly simple matter to show[9] that for sufficiently large N the sum in Eq. (2.14) is asymptotically equal to $(1+x)^N$, to first order in x. Therefore $\log K$ is is given by,

$$\log K=N\log V+N\log(1+x)\doteq N\log V+Nx. \tag{2.16}$$

This completes the evaluation of $\log K$ to the desired degree of approximation. We shall now show that a very similar cluster expansion can be obtained for $\log\langle R_{II}\rangle^0$.

The wave functions that appear in the integral, Eq. (2.10), for R_{II} are the correctly symmetrized wave functions for N noninteracting point particles of mass m'. They can be written as,

$$\phi_l(\mathbf{r}_1\cdots\mathbf{r}_N)=A_l\sum_P \exp[2\pi i P\sum_{j=1}^{N}\mathbf{k}_j\cdot\mathbf{r}_j], \tag{2.17}$$

where A_l is a normalization constant, P is a permutation operator acting on the N suffixes of the wave vectors $\mathbf{k}_j$, $\sum_P$ indicates a summation over all distinct permutations P and $\sum_j$ is a sum over all N indices j. From this last equation we have that,

$$\phi_l^*\phi_l=A_lA_l^*\sum_P\sum_{P'} \times\exp[2\pi i(P\sum_{j=1}^{N}\mathbf{k}_j\cdot\mathbf{r}_j-P'\sum_{j'=1}^{N}\mathbf{k}_{j'}\cdot\mathbf{r}_{j'})]. \tag{2.18}$$

Substituting this expression into Eq. (2.10) we have

$$R_{II}=A_lA_l^*\sum_P\sum_{P'}R_{II}(P,P'), \tag{2.19}$$

where

$$R_{II}(P,P')=\int\cdots\int\exp[2\pi i(P\sum_{j=1}^{N}\mathbf{k}_j\cdot\mathbf{r}_j-P'\sum_{j'=1}^{N}\mathbf{k}_{j'}\cdot\mathbf{r}_{j'})]\times\sum_{i<j}(1+f_{ij})d\mathbf{r}_1\cdots d\mathbf{r}_N. \tag{2.20}$$

We now expand the product $\prod_{i<j}(1+f_{ij})$ in the same way as before and we again select only those terms from the rth sum in the expansion that have no common suffices. This procedure yields the following series for $R_{II}(P,P')$,

$$R_{II}(P,P')=\sum_{r=0}^{N/2}P(r,N)N^{-r}J_{l,r}(P,P'), \tag{2.21}$$

where

$$J_{l,r}(P,P')=\int\cdots\int f_{ij}f_{kl}\cdots f_{\mu\sigma} \times\exp[2\pi i(P\sum_{j=1}^{N}\mathbf{k}_j\cdot\mathbf{r}_j-P'\sum_{j'=1}^{N}\mathbf{k}_{j'}\cdot\mathbf{r}_{j'})]. \tag{2.22}$$

We now substitute this series for $R_{II}(P,P')$ into Eq. (2.19) and get

$$R_{II}=A_lA_l^*\sum_P\sum_{P'}\sum_{r=0}^{N/2}P(r,N)N^{-r}J_{l,r}(P,P'). \tag{2.23}$$

If we invert the order of summation, this leads to

$$R_{II}=\sum_{r=0}^{N/2}P(r,N)N^{-r}H_{l,r}, \tag{2.24}$$

where $H_{l,r}$ is given by

$$H_{l,r}=A_lA_l^*\sum_P\sum_{P'}J_{l,r}(P,P'). \tag{2.25}$$

The evaluation of the integrals in Eq. (2.22) and the summation in Eq. (2.25) involve elementary but lengthy operations. These are carried out in Appendix A where it is shown that $H_{l,r}$ is given by

$$H_{l,r}=(I_l)^r=(N/V)^r\left[\int_{\rho} f(\boldsymbol{\rho})g_l(\boldsymbol{\rho})d\boldsymbol{\rho}\right]^r, \tag{2.26}$$

where

$$g_l(\boldsymbol{\rho})=(1/N^2)\sum_{\mathbf{k}\neq\mathbf{k}'}\sum n_{\mathbf{k}}n_{\mathbf{k}'}\exp[2\pi i\boldsymbol{\rho}\cdot(\mathbf{k}-\mathbf{k}')]+1, \tag{2.27}$$

and the symbols $n_{\mathbf{k}}$ that appears in this last equation are the set of integers that define the state l; they give the number of wave vectors in the wave function ϕ_l that are equal to $\mathbf{k}$, and satisfy the relation $\sum_{\mathbf{k}}n_{\mathbf{k}}=N$. We see that the only difference between the integral for I_l and the corresponding classical integral for x is the extra factor $g_l(\boldsymbol{\rho})$ in the integrand. This function, considered as a function of N and V, is of order unity.

We have at once from Eq. (2.24) that

$$\langle R_{11}\rangle^0=\sum_{r=0}^{N/2}P(r,N)N^{-r}\langle(I_1)^r\rangle^0. \tag{2.28}$$

It is shown in Appendix B that to our degree of approximation $\langle(I_1)^r\rangle^0$ can be replaced by $(\langle I_1\rangle^0)^r$; we then have

$$\langle R_{11}\rangle^0=\sum_{r=0}^{N/2}P(r,N)N^{-r}(\langle I_1\rangle^0)^r, \tag{2.29}$$

and for large N this gives,

$$\langle R_{11}\rangle^0\simeq(1+\tfrac{1}{2}\langle I\rangle^0)^N. \tag{2.30}$$

From Eq. (2.9), we have

$$F_1=-kT\log\langle R_{11}\rangle^0\simeq-\frac{NkT}{2}\langle I_1\rangle^0, \tag{2.31}$$

where

$$\langle I_1\rangle^0=(N/V)\int_{\varrho}f(\varrho)\langle g_1(\varrho)\rangle^0 d\varrho, \tag{2.32}$$

and

$$\langle g_1(\rho)\rangle^0=(1/N^2)\sum_{\mathbf{k}\neq\mathbf{k}'}\sum\bar{n}_{\mathbf{k}}\bar{n}_{\mathbf{k}'}\exp[2\pi i\rho\cdot(\mathbf{k}-\mathbf{k}')]+1. \tag{2.33}$$

In this last equation the $\bar{n}_{\mathbf{k}}$ are the mean values of the integers $n_{\mathbf{k}}$ that appeared in Eq. (2.26). Since $f(\varrho)$ differs from zero only if $|\varrho|\leqslant 2r_0$, we see that $\langle I_1\rangle^0\sim Nr_0^3/V$, and is therefore of exactly the same order of magnitude as the first term in the classical cluster expansion. It is easily shown that if we include in our calculations more complicated products of f_{ij}'s then we get terms in the expansion of F_1, that depend on higher powers of Nr_0^3/V. We shall not, however, calculate any of these terms as they are unnecessary for our purpose.

3. NATURE OF THE TRANSITION

If we combine Eq. (2.32) and Eq. (2.31), invert the order of integration and summation and then substitute the result in Eq. (2.30) we get,

$$F_1=\tfrac{1}{2}\sum_{\mathbf{k}\neq\mathbf{k}'}\sum\bar{n}_{\mathbf{k}}\bar{n}_{\mathbf{k}'}L_{\mathbf{k},\mathbf{k}'}+\tfrac{1}{2}N^2L_0, \tag{3.1}$$

where

$$L_{\mathbf{k},\mathbf{k}'}=\frac{1}{V}\int_{\varrho}\exp[2\pi i\varrho\cdot(\mathbf{k}-\mathbf{k}')][-kTf(\varrho)]d\varrho, \tag{3.2}$$

and

$$L_0=\frac{1}{V}\int_{\varrho}[-kTf(\varrho)]d\varrho. \tag{3.3}$$

Now the $\bar{n}_k$ that occur in Eq. (3.1) are, as we have pointed out in the previous section, the average values of the integers n_k that define the state l. Since these averages were calculated with weights $\exp(-\beta E_l)/\sum\exp(-\beta E_l)$, the $\bar{n}_k$ must be identical with the average occupation numbers of the single particle energy levels of an ideal Bose-Einstein gas. We summarize the properties[9] of these occupation numbers,

$$\bar{n}_{\mathbf{k}}=[\exp(\mu k^2+\alpha)-1]^{-1},\quad \text{all } \mathbf{k};\quad T>T_\lambda, \tag{3.4}$$

$$\bar{n}_{\mathbf{k}}=[\exp(\mu k^2)-1]^{-1},\quad \mathbf{k}>0;\quad T\leqslant T_\lambda, \tag{3.5}$$

and

$$\bar{n}_0=N-\sum_{\mathbf{k}\neq 0}\bar{n}_{\mathbf{k}}.\quad T\leqslant T_\lambda$$

Here $\mu=h^2/2m'kT$, the transition temperature T_λ is given by

$$T_\lambda=(h^2/2m'k\pi)[N/V\zeta(\tfrac{3}{2})]^{\frac{2}{3}}, \tag{3.6}$$

and $\zeta(\frac{3}{2})$ is the Riemann zerta function of order $\frac{3}{2}$. The statistical parameter α is given as a function of T and N/V by the equation $\sum_{\mathbf{k}}\bar{n}_{\mathbf{k}}(\alpha;T,V)=N$, for $T>T_\lambda$. For $T\leqslant T_\lambda$ α is identically zero.[9] It follows from this last equation that α and its first order derivatives with respect to T and V are continuous at T_λ but that all the higher order derivatives are discontinuous at this temperature. Equations (3.4) and (3.5) define the $\bar{n}_{\mathbf{k}}$ as functions of T, V, and α, and we shall henceforth consider F_1 to be given as a function of these variables. The term F_0 in Eq. (2.7) is very easily evaluated as it is the free energy of an ideal Bose-Einstein gas composed of particles of mass m'. It is given explicitly by the equations

$$F_0=-NkT\alpha+kT\sum_{\mathbf{k}}\log[1-\exp(-\mu k^2-\alpha)];\quad T>T_\lambda, \tag{3.7}$$

and

$$F_0=kT\sum_{\mathbf{k}}\log[1-\exp(-\mu k^2)];\quad T\leqslant T_\lambda. \tag{3.8}$$

Here all the symbols have exactly the same meaning as before. If we regard Eq. (3.7) as defining F_0 as an explicit function of T, V, and α, then it is easily seen that $[\partial F_0/\partial\alpha]_{T,V}\equiv 0$; this last relation is in fact equivalent to the equation $\sum_{\mathbf{k}}\bar{n}_k(\alpha,T,V)=N$, which the $\bar{n}_{\mathbf{k}}$ must satisfy.

We can now use exactly the same arguments that we used in a previous publication[10] to show that, while F_0 exhibits a third-order transition at T_λ, F_1, exhibits a second order transition at this temperature. We first notice that because $[\partial F_0/\partial\alpha]_{T,V}\equiv 0$, the rth derivatives of F_0 with respect to T and V only depend on the $(r-1)$th and lower order derivatives of α. Since only the second and higher order derivatives of α are discontinuous at T_λ it follows that only the third and higher order derivatives of F_0 will be discontinuous at this temperature. This result is of course to be expected because F_0 is merely the free energy of an ideal Bose-Einstein gas, and such a system exhibits a third order transition under conditions of constant volume. It is easily seen from Eq. (3.1) that $[\partial F_1/\partial\alpha]_{T,V}\neq 0$. Therefore the rth derivatives of F_1 will depend on the rth derivatives of α. This in turn implies that the second-order derivatives of F_1 will be discontinuous at

[9] A. R. Fraser, Phil. Mag. **42**, 156, 165 (1951).
[10] G. V. Chester, Phys. Rev. **94**, 246 (1954).

T_λ and that F_1 will exhibit a second-order transition under conditions of constant volume. To show that it exhibits a second-order transition under conditions of constant pressure we merely have to show that $(\partial p/\partial V)_T<0$, for $T\leqslant T_\lambda$. This inequality also determines whether the system is stable for $T\leqslant T_\lambda$. Because $(\partial p/\partial V)_T\equiv 0$, for $T\leqslant T_\lambda$, in the London theory this theory leads to a first-order transition under conditions of constant pressure. It is easily shown[11] that the first term in Eq. (3.1) is linear in V, for $\leqslant T_\lambda$; therefore $(\partial p/\partial V)_T$ is given by[12]

$$(\partial p/\partial V)_T=(-N/V^2)\int_{\varrho}[-kTf(\varrho)]d\varrho. \quad (3.9)$$

Consequently $(\partial p/\partial V)_T<0$ as long as $\int_{\varrho}[-kTf(\varrho)]d\varrho>0$. From the definition of $f(\varrho)$, Eq. (1.3), we have at once the integral $=4\pi MkTr_0^3/3V^2$, and is therefore positive. We have therefore shown that to our degree of approximation the partition function proposed by Feynman exhibits a second-order transition both under conditions of constant pressure and under conditions of constant volume. For the sake of completeness we quote the formula for the discontinuity in the specific heat at constant volume. If $\Delta C_V=C_V(T_\lambda+0)-C_V(T_\lambda-0)$, then

$$\Delta C_V=4\pi N\frac{N}{V}\frac{T_\lambda^2}{\zeta^2}\left(\frac{\partial^2\alpha}{\partial T^2}\right)_{T=T_\lambda}\int_0^\infty[-kTf(\varrho)\times S(\rho)T(\varrho)]\rho^2d\rho, \quad (3.10)$$

where

$$S(\rho)=\sum_{n=1}^{\infty}\exp(-a\rho^2/n)/n^{\frac{1}{2}},$$

$$T(\rho)=\sum_{n=1}^{\infty}[1-\exp(-a\rho^2/n)]/n^{\frac{1}{2}}$$

and $a=\pi/\lambda'$.

It is interesting to examine, in a more physical manner, the reason why the second term in the expansion changed the transition from one of the third order to one of the second order. In the first term in the series the transition arises because of the existence of very large cyclical configurations, or polygons, in the integrand of the partition function. This fact has been pointed out by Feynmann.[1] Now it is well known that in the type of cluster expansion we have used the second term in the expansion always introduces correlation between the atoms. In particular this means that we have introduced correlation between the polygons in the integrand of the partition function. It is just the introduction of these correlations that changes the order of the transition. As long as the polygons are uncorrelated we shall have a third-order transition, but as soon as we introduce correlation between them the transition changes to a second-order one.

Finally we note that if T_λ, as given by Eq. (3.6), is to agree with experiment we must have then $m'\sim\frac{1}{2}m$ at the λ temperature. We also note that $\partial T_\lambda/\partial\rho$ will only have the correct sign if m' varies with ρ faster than $\rho^{\frac{1}{2}}$. While both these requirements seem to be reasonable on physical grounds we cannot see any rigorous method of showing that they are satisfied.

We now turn to a discussion of the validity of these results, the point at issue being whether they are true for the exact form of Feynman's partition function or whether they are spurious and result from the approximations that we have made. The two essential steps in our method are (a) the development of the expansion formula itself and (b) the approximation of breaking off this expansion after only two terms. We shall discuss these points in turn.

The use of an expansion formula for F_1 could only lead to incorrect results if it were divergent. To develop the complete series for F_1 along the lines we have suggested would be an extremely tedious task and a rigorous proof, or disproof, of the convergence of the series would be almost impossible. It would therefore seem to be worth while to present a much simpler physical argument that suggests that the series for F_1 is convergent at the temperatures and densities we are interested in. Our argument is based on the fact that if we set $\phi_l\phi_l^*\neq 1$ in the integral for R_{ll} then this integral becomes identical with the classical configuration integral for N perfectly elastic spheres; our expansion formula is then identical with the classical virial expansion. Now since the coefficients of this expansion are independent of the temperature its radius of convergence will also be independent of the temperature. It seems reasonable on physical grounds to suppose that this series will be convergent for sufficiently low densities, and, since the density of liquid helium is only about one half the density of a "normal" fluid, it is reasonable to suppose that the series will be convergent at liquid helium densities. Now the only difference between the coefficients in the classical expansion and those in the expansion we have proposed is that in the integrals that define the coefficients we have an extra factor $\phi_l\phi_l^*$. This factor is always positive, less than unity and, in the region we are interested in, is a slowly varying function of the coordinates. Bearing these facts in mind it seems possible that the introduction of this extra factor into the integrals will in fact leave the radius of convergence of the series almost unaltered. On the basis of this very incomplete and unrigorous argument it is plausible that the expansion we have proposed for F_1 is convergent at the temperatures and densities we are interested in.

If this conclusion is correct then we only have to show that the results we have obtained on the basis of

[11] G. V. Chester, Phys. Rev. **100**, 446 (1955).

[12] This formula is correct as long as the dependence of m' on V is ignored. If this dependence is taken into account then the expression for $(\partial p/\partial V)_T$ becomes more complicated. It is easily seen, however, that $(\partial p/\partial V)_T\neq 0$ and that in consequence the transition remains a second order one.

the first two terms in the series will not be altered by taking into account the behavior of any of the higher terms in it. We in fact showed that the second order derivatives of the second term in the expansion of F_1 were discontinuous at a temperature T_λ given by Eq. (3.6). If this result is spurious then the derivatives of the higher order terms in the series must also be discontinuous at this temperature. Moreover these discontinuities must be just sufficient to cancel out those that we have found. It is very easily seen that our expansion of F_1 is essentially an expansion in powers of Nr_0^3/V. Consequently while the discontinuities in the various terms may exactly cancel for one particular value of Nr_0^3/V it would seem to be impossible that they should do so in general. We can therefore conclude that the discontinuities we have found are in fact genuine. On the other hand it should be pointed out that there is nothing to prevent the first-order derivatives of any of the higher order terms in the series exhibiting discontinuities; this type of behavior would lead to a first order transition. We can therefore conclude that while it seems fairly certain that Feynman's partition function exhibits a second-order λ transition we cannot rule out the possibility that it exhibits a first-order λ transition.

4. SOME GENERAL CONSIDERATIONS

First we wish to draw attention to an important theoretical point that must be considered whenever we construct a theory of liquid helium. A helium atom consists of two electrons that are tightly bound to a nucleus, and we normally treat this composite system as an individual particle. Moreover since the total spin of this system, when it is in the ground state, is zero we must apply Bose-Einstein statistics to any assembly of them. However, this procedure is only correct as long as the two electrons can be considered as being so tightly bound to the nucleus that the whole system behaves as a rigid particle. This is of course well known.[13] Now if in any statistical treatment of a system composed of helium atoms we allow configurations to occur in which two or more atoms come so close together that their charge clouds overlap, then we cannot apply Bose-Einstein statistics to the atoms in these configurations. In other words, in configurations of this type we must not symmetrize the total wave function with respect to the coordinates of the atoms but we must rather antisymmetrize it with respect to the coordinates of the individual electrons and nucleons that go to make up the atoms. For this reason, it is not possible to regard London's ideal gas theory as an accurate treatment of a noninteracting system of helium atoms, although it is of course an accurate treatment of a system of noninteracting Bose-Einstein particles. For example, it is clear from the Pauli exclusion principle that the probability of finding two collections of electrons and nuclei in the same region of configuration space will be very small. London's treatment, on the other hand, indicates that there is a very high probability of this happening.

Let us now examine Feynman's partition function in the light of these remarks. It follows from Eq. (1.1) that Feynman gives each configuration a statistical weight $\rho_0 \exp(-\beta V')$. Therefore as long as the function $\exp(-\beta V')$ vanishes whenever two or more helium atoms approach very close together this configuration density will be perfectly satisfactory. To be more precise this means that the function $\exp(-\beta V')$ must vanish, or at least become very small, whenever any of the $N(N-1)/2$ conditions $r_{ij} \lesssim 2r_0$ (where r_0 is the radius of the charge cloud of a helium atom) are satisfied. However, we have pointed out that this is an essential property of the function $\exp(-\beta V')$. We can therefore conclude that the configurational density that Feynman has proposed fulfills the requirements we have discussed above. It is easily shown that our approximate method of dealing with the partition function is also satisfactory in this respect.

Finally we wish to discuss whether Feynman's partition function can exhibit a liquid-gas transition. This question is important because the λ transition only takes place in the liquid phase of helium. Consequently any really satisfactory theory of liquid helium must (a) show that helium exhibits a liquid-gas transition and (b) show that this transition takes place at a temperature above the theoretical λ transition. It appears likely that Feynman's treatment can fulfill at least the first of these requirements. This conclusion rests on the following considerations. The function $\rho(\mathbf{r}_1 \cdots \mathbf{r}_N)$, that occurs in the integrand of the partition function, is of the form $1+f(\mathbf{r}_1 \cdots \mathbf{r}_N)$. Here $f(\mathbf{r}_1 \cdots \mathbf{r}_N)$ is a sum of terms each of which contains at least one factor of the type $\exp[-(8\pi r_0^2/\lambda^2)x_{ij}^2]$, where $x_{ij}=r_{ij}/2r_0$ and $\lambda=h^2/2\pi mkT$. Since the other factor in the integrand vanishes whenever $x_{ij} \lesssim 1$ the smallest value that the argument of the exponential can have is $8\pi r_0^2/\lambda^2$. This in turn implies that the function $f(\mathbf{r}_1 \cdots \mathbf{r}_N)$ falls very rapidly to zero, as any pair of relative coordinates increases, as long as $T \gg 2°\mathrm{K}$. Thus for temperatures above about 6°K the factor ρ_0 may be replaced by unity. In this region of temperature therefore the partition function reduces to exactly the same form as the partition function for a classical fluid except that the actual potential V is replaced by the quasi potential $V'(\beta)$. Now as long as $V'(\beta)$ contains a weak attractive part this partition function will exhibit a gas-liquid transition. Unfortunately we cannot estimate the temperature at which this will occur unless we know something of the form of $V'(\beta)$, and this means that we cannot be sure that even if a transition takes place it will occur at a temperature above the λ point. However, it is worth noting that if V' is set equal to the true interaction potential, V, then it is almost certain that the partition function will exhibit a gas-liquid transition at about 13°K. For we know from the law of

[13] R. H. Fowler, reference 8, Chap. 2.

corresponding states that if helium were a "classical" liquid it would liquify at about this temperature. Now at 13°K the factors $\exp(-kr^2)$ that occur in the function $f(\mathbf{r}_1\cdots\mathbf{r}_N)$ are never greater than 10^{-3}, in the region of configuration space of interest, and hence we can certainly set $f\equiv 0$ and $\rho_0\equiv 1$. The partition function then becomes identical with that for a classical fluid, with an interaction potential V, and will therefore exhibit a gas-liquid transition at about 13°K. This shows that this approximation to the partition function does not take into account the high zero-point energy of liquid helium. To introduce the zero-point energy in this approximation we have to assume that the quasi potential, V', is much weaker than the actual potential V. On the basis of this assumption it is probably fairly simple to construct a realistic "model" of liquid helium near the λ point; it remains to be shown how such a "model" can be derived from first principles. As we have already mentioned, the weak attractive part of the potential, that is necessary for a gas-liquid condensation, can easily be taken account of near the λ temperature by means of perturbation theory.

The author wishes to thank the United States Government for the award of an F. O. A. research fellowship.

APPENDIX A

In this Appendix, we shall outline the main steps in the evaluation of the integral for $J_{L,r}(P,P')$, Eq. (2.22), and the sum for $H_{L,r}$, Eq. (2.25).

We first consider the integral for $J_{L,s}(P,P')$,

$$J_{L,s}(P,P')=\int\cdots\int f(1,2)f(3,4)\cdots f(2s-1,2s)\times\exp[2\pi i(P\sum_{j=1}^{N}\mathbf{k}_j\cdot\mathbf{r}_j-P'\sum_{j'=1}^{N}\mathbf{k}_{j'}\cdot\mathbf{r}_{j'})]d\mathbf{r}_1\cdots d\mathbf{r}_N, \quad (A.1)$$

where $f(i,k)$ stands for $f(r_{ik})$. Let us suppose that P and P' are a pair of permutations that permute the indices (j) on the wave vectors k in the following way,

$$P(1\cdots N)=(\alpha_1\cdots\alpha_N), \quad (A.2)$$

and

$$P'(1\cdots N)=(\alpha_1'\cdots\alpha_N'). \quad (A.3)$$

These two equations are to be interpreted as meaning that when P operates on the indices $1\cdots N$ it sends them into a permutation $\alpha_1\cdots\alpha_N$ and similarly for P'. There are $(N-2r)$ coordinates that do not appear in the functions $f(i,k)$. We can therefore integrate over these coordinates at once. Because of the orthogonality conditions on the single particle wave functions, $\exp[2\pi i\mathbf{k}\cdot\mathbf{r}]$, the integration over the ith of these coordinates yields a factor $V\delta(\alpha_i,\alpha_i')$, where $\delta(\alpha,\beta)$ is zero if $\alpha\neq\beta$ and unity if $\alpha=\beta$. Consequently, $J_{L,s}(P,P')$ will vanish unless $\alpha_i=\alpha_i'$, for all $i>2s$. When these conditions are fulfilled the integral reduces to,

$$J_{L,s}(P,P')=V^{N-2s}\int\cdots\int f(1,2)f(3,4)\cdots f(2s-1,2s)\times\exp[2\pi i(P\sum_{j=1}^{2s}\mathbf{k}_j\cdot\mathbf{r}_j-P'\sum_{j'=1}^{2s}\mathbf{k}_{j'}\cdot\mathbf{r}_{j'})]d\mathbf{r}_1\cdots d\mathbf{r}_{2s}. \quad (A.4)$$

Next we introduce center of mass and relative coordinates for each pair of coordinates that appear together in a function $f(i,i+1)$. Thus,

$$\boldsymbol{\rho}_i=(\mathbf{r}_i-\mathbf{r}_{i+1})/2;\quad \mathbf{R}_i=(\mathbf{r}_i+\mathbf{r}_{i+1})/2;\quad (i=1,3,\cdots,2s-1). \quad (A.5)$$

Solving these equations we get,

$$\mathbf{r}_i=\mathbf{R}_i+\boldsymbol{\rho}_i;\quad \mathbf{r}_{i+1}=\mathbf{R}_i-\boldsymbol{\rho}_i;\quad (i=1,3\cdots,2s-1), \quad (A.6)$$

and the integral now becomes

$$J_{L,s}(P,P')=V^{N-2s}\int\cdots\int f(\boldsymbol{\rho}_1)\cdots f(\boldsymbol{\rho}_{2s-1})\times\exp\{2\pi i\sum_{i=1,3\cdots}^{2s-1}\mathbf{R}_i[P(\mathbf{k}_i+\mathbf{k}_{i+1})-P'(\mathbf{k}_i+\mathbf{k}_{i+1})]\}\times\exp\{2\pi i\sum_{i=1,3\cdots}^{2s-1}\boldsymbol{\rho}_i[P(\mathbf{k}_i-\mathbf{k}_{i+1})-P'(\mathbf{k}_i-\mathbf{k}_{i+1})]\}\times d\boldsymbol{\rho}_1\cdots d\boldsymbol{\rho}_{2s-1}d\mathbf{R}_1\cdots d\mathbf{R}_{2s-1}. \quad (A.7)$$

If we perform the integrations over the center-of-mass coordinates $\mathbf{R}_j$, we find that $J_{L,s}(P,P')$ vanishes unless the s conditions,

$$\mathbf{k}(\alpha_i)+\mathbf{k}(\alpha_{i+1})=\mathbf{k}(\alpha_i')+\mathbf{k}(\alpha_{i+1}'),\quad (i=1,3\cdots,2s-1). \quad (A.8)$$

are satisfied. Here $\mathbf{k}(\mu)$ stands for $\mathbf{k}_\mu$. These conditions can be fulfilled in two ways: (a) by choosing P and P' so that they are satisfied identically for all values of the wave vectors $\mathbf{k}$ or (b) by regarding them as conditions to be imposed on the wave vectors. We ignore this second possibility for the moment; we shall show later that if one or more of the s conditions are imposed on the wave vectors then the contribution to $H_{L,s}$ is negligible. Let us return to the first possibility. For any given P the conditions given by Eq. (A.8) can only be satisfied if P' is such that either $\alpha_i'=\alpha_i$ and $\alpha_{i+1}'=\alpha_{i+1}$, or $\alpha_i'=\alpha_{i+1}$ and $\alpha_{i+1}'=\alpha_i$, for all $i=1,3\cdots,2s-1$. Let us suppose therefore that P' is a permutation such that $\alpha_i=\alpha_{i+1}'$ and $\alpha_{i+1}=\alpha_i'$, for $1\leqslant i\leqslant p$; while $\alpha_{i+1}'=\alpha_{i+1}$ and $\alpha_i'=\alpha_i$, for $p<i\leqslant 2s-1$. Then the integral for

$J_{L,s}(P,P')$ becomes

$$J_{L,s}(P,P')=J_{L,s}(P,\hat{p})$$

$$=V^{N-s}\left[\int_{\varrho} f(\varrho)d\varrho\right]^{s-p}\int\cdots\int f(\varrho_1)\cdots$$

$$\times f(\varrho_p)\exp\left[2\pi i\sum_{i=1}^{p}\varrho_i\cdot(\mathbf{k}_i-\mathbf{k}_{i+1})\right]d\varrho_1\cdots d\varrho_p,$$

where we have relabeled the wave vectors for convenience, $k_i=k(\alpha_i)$, and we have replaced P' in $J_{L,s}(P,P')$ by $\hat{p}$ to indicate that P' is identical with P except that p of the N indices on the k's are permuted in pairs, namely so that $\alpha_i'=\alpha_{i+1}$ and $\alpha_{i+1}'=\alpha_i$, for $1\leqslant i\leqslant p$. This last equation is the complete expression for $J_{L,s}(P,P')$ when P is any permutation and P' differs from P in the way we have specified. If we were to choose a more complicated pair of permutations that could not be specified in this way then we could not satisfy the s conditions (A.8) on the wave vectors identically.

We next consider the sum for $H_{L,s}$,

$$H_{L,s}=A_L A_l^{*}\sum_{P}\sum_{P'}J_{L,s}(P,P'). \tag{A.10}$$

Now since the permutation P' is labeled by the number of pairs of adjacent indices that it permutes we have at once that

$$\sum_{P'}J_{L,s}(P,P')=\sum_{p=0}^{s}{}^{s}C_p J_{L,s}(P,\hat{p}). \tag{A.11}$$

The permutation P is on the other hand labeled by the particular wave vectors that are associated with the coordinates $\mathbf{r}_j$. But $J_{L,s}(P,\hat{p})$ depends only on the pairs of wave vectors that are associated with relative coordinates $\varrho_1\cdots\varrho_p$. Therefore all permutations that permute the N indices so that the same wave vectors $\mathbf{k}$ are associated with the coordinates $\varrho_1\cdots\varrho_p$ will lead to the above expression for $J_{L,s}(P,\hat{p})$. Since we have $(N-2p)$ wave vectors that do not appear in the expression $J_{L,s}(P,\hat{p})$ and these vectors are equal in groups $n_{\mathbf{k}_i}-1$ or $n_{\mathbf{k}_i}$, according as one of the vectors in the group $n_{\mathbf{k}_i}$ appears in $J_{L,s}$ or not, the number of permutations P that lead to the same expression for $J_{L,s}$ is simply

$$(N-2p)!/\prod_{i=1}^{2p}(n_{\mathbf{k}_i}-i)!\prod_{j>2p}(n_{\mathbf{k}_j})!. \tag{A.12}$$

To complete the summation over P we merely have to sum over all values of the wave vectors in $J_{L,s}$. After a few elementary reductions we get,

$$H_{L,s}=\sum_{p=0}^{s}{}^{s}C_p V^{s-p}K^{s-p}L_l{}^{p}(N-2_p)!/N!, \tag{A.13}$$

where we have used the fact that

$$A_l A_l{}^{*}=V^{-N}\prod_{j}(n_{\mathbf{k}_j})!/N!,$$

$$K=\int_{\varrho} f(\varrho)d\varrho, \tag{A.16}$$

and

$$L_l=\sum_{\mathbf{k}_1\neq\mathbf{k}_2}\sum n_{\mathbf{k}_1}n_{\mathbf{k}_2}\int_{\varrho} f(\varrho)\exp[2\pi i\varrho\cdot(\mathbf{k}_1-\mathbf{k}_2)]d\varrho. \tag{A.15}$$

Consequently, for large N,

$$H_{L,s}=(N/V)^{s}[K+L_l/N^2]^{s}=(I_l)^{s}, \tag{A.16}$$

and this is the result quoted in the text.

We can now easily see why any pairs of permutations that are such that not all the s conditions (A.8) are fulfilled identically do not contribute significantly to $H_{L,s}$. For if t of these conditions are not satisfied identically then they must be imposed as restrictions on the values that the wave vectors $\mathbf{k}$ can take. Therefore the contribution to $H_{L,s}$ from a permutation of this type will be exactly the same as that given by Eq. (A.16) except that there will be t restrictions on the summations over the $\mathbf{k}$'s. Now since the $n_{\mathbf{k}}$ obey the condition $\sum_{\mathbf{k}} n_{\mathbf{k}}=N$ each restriction we impose on the summations effectively introduces a factor N^{-1} in $H_{L,s}$. If we have t restrictions then we introduce a factor N^{-t} and the contribution to $H_{L,s}$ will therefore be negligible for large N.

APPENDIX B

In this Appendix, we shall show that it is permissible to replace $\langle(I_l)^r\rangle^0$ by $(\langle I_l\rangle^0)^r$ in Eq. (2.27). First we notice that $\langle(I_l)^r\rangle^0=(\langle I_l\rangle^0)^r$, if $r=0$ or 1. Consequently the error introduced into F_1 by this approximation can only depend on second and higher powers of the integral I_l. Let us now write

$$\langle(I_l)^r\rangle^0=(\langle I_l\rangle^0)^r(1+\Delta_r). \tag{B.1}$$

Then it can be easily shown, using the methods developed by Fowler,[14] that

$$|\Delta_r|\leqslant(1+br/N)^r-1\ll(1+b)^r, \tag{B.2}$$

where b is a quantity of order unity. From Eqs. (B.1) and (B.2) together with Eqs. (2.29) and (2.30) of the text, we see that the magnitude of the error, ΔF_l, in F_1 obeys the following inequality

$$|\Delta F_1|\ll N|\langle I_l\rangle^0 bkT|, \tag{B.3}$$

and is therefore at most of order N. If we now bear in mind that ΔF_1 can only depend on second and higher powers of I_l then we see that it can be written in the form $Nf(I_l)$, where $f(I_l)$ depends only on second and higher powers of I_l. But any quantity of this order is negligible compared with $N\langle I_l\rangle^0$ and therefore the error introduced into F_1 by this approximation is negligible.

[14] R. H. Fowler, reference 8, Chap. 20.

Superconductivity of a Charged Ideal Bose Gas

M. R. SCHAFROTH
The F. B. S. Falkiner Nuclear Research and Adolph Basser Computing Laboratories, School of Physics, The University of Sydney, Sydney, Australia*
(Received April 4, 1955)

It is shown that an ideal gas of charged bosons exhibits the essential equilibrium features of a superconductor. The onset of Bose-Einstein condensation marks the transition temperature T_c. Below T_c a Meissner-Ochsenfeld effect is exhibited which is described in a very good approximation by London's equation.

The singular nature of the condensed ideal Bose gas exhibits itself in a space dependence of the London constant λ, determined by the boundary conditions on the wave function. It is shown that the electrostatic repulsion between the bosons compensates this effect and leads to a spatially constant λ, independently of the boundary conditions.

The critical field $H_c(T)$ is determined and found to be related to the penetration depth $d(T)$ by

$$H_c=\hbar c/2ed^2$$

(e being the boson charge).

The $B(H)$ law is different from the one usually assumed for actual superconductors. Corresponding changes occur in the thermodynamical relation.

A comparison with superconducting metals is made. The main conclusion is that if superconductivity in metals is due to the concurrence of bosons, then the number of these bosons must be strongly temperature-dependent below T_c.

1. INTRODUCTION

So far, no molecular theory of superconductivity has been found. The most successful attempts in this direction have been made by Froehlich and Bardeen[1] on the assumption that the occurrence of this phenomenon is due to the interaction of the conduction electrons with lattice vibrations. This assumption led to reasonable estimates of the energy values involved and was thus able to explain the isotope effect.[2] However, it has so far not been possible to show that a strong enough lattice-electron interaction can account for the characteristic equilibrium phenomena of superconductivity, namely the phase transition and the Meissner-Ochsenfeld effect. In the weak-coupling approximation employed by Froehlich[1] these effects certainly do not occur.[3]

The effects of the lattice-electron interaction in any other than the weak-coupling approximation are very complex, and it is difficult to foresee which will be the characteristic feature responsible for a transition to the superconducting state. This holds especially because until recently no simple physical model was known which exhibits the equilibrium properties of a superconductor and which could serve as a lead in the search for the real phenomenon.

Such a model has recently been pointed out by the author[4]: it is the ideal gas of charged bosons. The purpose of this paper is to prove this assertion and to discuss the detailed properties of this model.

The fact that an ideal Bose gas exhibits a thermodynamic transition point of the second kind[5] is well known[6] and need not be dealt with here. The transition temperature T_c is given by

$$kT_c=4\pi\frac{\hbar^2}{2m}\left[\frac{n}{\zeta(\frac{3}{2})}\right]^{\frac{2}{3}}, \tag{1.1}$$

where m is the mass of the bosons, n their density, k the Boltzmann constant and

$$\zeta(s)=\sum_{n=1}^{\infty}n^{-s} \quad (s>1) \tag{1.2}$$

is the Riemann ζ function.

In contrast to real superconductors, the specific heat is continuous through the transition temperature, and only its derivative exhibits a jump. (Such a transition is often called "of the third kind.")

At temperature $T<T_c$ a finite fraction n_s of the total boson density n is condensed in the ground state, the remaining part $n_n=n-n_s$ forms a normal Bose gas with chemical potential $\mu=0$.[7] The densities in the two phases are given by

$$n_n=n(T/T_c)^{\frac{3}{2}}, \tag{1.3}$$

$$n_s=n[1-(T/T_c)^{\frac{3}{2}}]. \tag{1.4}$$

An indication that at the condensation point an ideal charged Bose gas becomes superconducting is found in computing its magnetic susceptibility. The magnetic

* Also supported by the Nuclear Research Foundation within the University of Sydney.

[1] H. Froehlich, Phys. Rev. **79**, 845 (1950); J. Bardeen, Revs. Modern Phys. **23**, 261 (1951), and references given there.

[2] E. Maxwell, Phys. Rev. **78**, 477 (1950); **79**, 173 (1950); Reynolds, Serin, Wright, and Nesbitt, Phys. Rev. **78**, 487 (1950); Serin, Reynolds, and Nesbitt, Phys. Rev. **78**, 813 (1950).

[3] M. R. Schafroth, Helv. Phys. Acta **24**, 645 (1951).

[4] M. R. Schafroth, Phys. Rev. **96**, 1149 (1954).

[5] We call any transition without latent heat "of the second kind."

[6] A. Einstein, Ber. Berl. Akad. 261 (1924); 3 (1925); F. London, Phys. Rev. **54**, 947 (1938); B. Kahn and G. E. Uhlenbeck, Physica **5**, 399 (1938).

[7] These statements hold in the limit of infinitely large volume G, i.e., neglecting G^{-1} compared to unity. For finite volumes, μ is of the order G^{-1}, depending on the shape of the volume, and there is no sharp transition.

susceptibility is defined by

$$M=\chi H,$$

where H is the applied field; what is actually computed, however, is a quantity χ' defined by

$$M=\chi' H',$$

where H' is the "acting" field. The relation between H and H' requires special attention; this point will be discussed in Sec. 5.

For an ideal charged Bose gas, χ' tends to $-\infty$ on approaching condensation. Quite generally, for an ideal gas, χ' is given by[8]

$$\chi'=-\tfrac{1}{6}\mu_0^2 n\langle E^{-1}\rangle_{kin}, \tag{1.5}$$

where $\mu_0=eh/2mc$ is Bohr's magneton $n=N/V$ the density of particles, and $\langle E^{-1}\rangle_{kin}$ is the average of the reciprocal kinetic energy of a particle over the distribution function. For a Bose gas this is

$$N\langle E^{-1}\rangle_{kin}=\frac{(2m)^{\frac{1}{2}}}{4\pi^2\hbar^3}\int_0^\infty d\epsilon\,\epsilon^{-\frac{1}{2}}\{\exp[\alpha(\epsilon-\mu)]-1\}^{-1} \tag{1.6}$$

[$\alpha=1/kT$; $\mu=$ chemical potential ($\mu<0$)]. For small $|\mu|$ this is of the order $|\mu|^{-\frac{1}{2}}$; on approaching condensation, $\mu\to 0$ and therefore $\chi'\to\infty$. (For a finite volume, μ and χ' become volume- and shape-dependent, and (1.5) is no longer valid.)

To establish the Meissner-Ochsenfeld effect, however more is needed. The current density induced by an *inhomogeneous* magnetic field has to be studied. The ensuing relation between current and field together with Maxwell's equation

$$\operatorname{curl}\mathbf{H}=(4\pi/c)\mathbf{i} \tag{1.7}$$

must then lead to an expulsion of the field from the system.

It is advantageous to expand the magnetic field in terms of a complete orthonormal set of functions appropriate to the shape of the volume under study. For a cubical box of volume G with periodical boundary conditions, which is the easiest volume to work with, these eigenfunctions are plane waves

$$\psi_{\mathbf{q}}(\mathbf{x})=G^{-\frac{1}{2}}\exp(i\mathbf{q}\cdot\mathbf{x}). \tag{1.8}$$

We restrict ourselves to weak fields, so that the relation between field and induced current can be taken as linear. In view of the translational symmetry of the particular volume chosen, this relation between current density $\mathbf{i}(\mathbf{x})$ and vector potential $\mathbf{A}(\mathbf{x})$ takes the simple general form

$$i_\mu(\mathbf{x})=\sum_{\nu=1}^{3}\int_G d^3\mathbf{x}' K_{\mu\nu}(\mathbf{x}-\mathbf{x}')A_\nu(\mathbf{x}'). \tag{1.9}$$

[8] See reference 3, appendix; see also Sec. 2 of the present paper.

In terms of the Fourier transforms [i.e., the coefficients in an expansion in terms of (1.8)], this becomes

$$i_\mu(\mathbf{q})=\sum_{\nu=1}^{3}K_{\mu\nu}(\mathbf{q})A_\nu(\mathbf{q}). \tag{1.10}$$

Gauge invariance requires[3]

$$K_{\mu\nu}(\mathbf{q})=(q_\mu q_\nu-\delta_{\mu\nu}q^2)K(q^2). \tag{1.11}$$

Normal diamagnetism corresponds to a $K(q^2)$ regular for small values of q:

$$K(q^2)=-c\chi'+O(q^2) \tag{1.12}$$

(c is the velocity of light). A pole at $q=0$ gives a Meissner-Ochsenfeld effect; in particular, the form

$$K(q^2)=1/\lambda c q^2 \tag{1.13}$$

is equivalent to the London[9] equation

$$-\lambda c\,\operatorname{curl}\mathbf{i}=\mathbf{H}. \tag{1.14}$$

The main result of this paper is that the ideal Bose gas below its condensation point obeys London's equation (1.14) with

$$\lambda^{-1}=(e^2/m)n_s, \tag{1.15}$$

where n_s is the density of condensed particles (1.4). This holds except for additional terms in $K(q^2)$ proportional to $|\mathbf{q}|^{-1}$ which are, however, practically negligible in all cases.

Special attention has to be devoted to the influence of the boundary condition on the boson wave functions. The calculations in Secs. 2 and 3 are devoted to the simplest case of a cube with periodical boundary condition. As, however, the occurrence of the Meissner-Ochsenfeld effect is due to the condensed bosons, i.e., to a finite fraction of the total particle density occupying a state which extends over the whole volume, there is no reason to assume that the relation between current density and field is independent of the boundary condition. A simplified calculation is carried through in Sec. 4; it takes into account only the contribution of the condensed bosons, relying on the discussion in Sec. 3 that the noncondensed particles do not contribute significantly. The result is that indeed the London equation (1.14) has to be replaced in general by

$$-c\,\operatorname{curl}(\lambda\mathbf{i})=\mathbf{H}, \tag{1.14'}$$

with

$$\lambda^{-1}=(e^2/m)n_s(\mathbf{x}), \tag{1.15'}$$

where now $n_s(\mathbf{x})$, the local density of bosons, is in general a function of position, depending upon the boundary condition. For a cube of volume $G=L^3$ with perfectly reflecting walls,

$$n_s(x)=n_s^0\cdot\frac{8}{G}\sin^2\frac{\pi x}{L}\sin^2\frac{\pi y}{L}\sin^2\frac{\pi z}{L}, \tag{1.16}$$

[9] F. London, *Superfluids I* (John Wiley and Sons, Inc., New York, 1950).

where n_s^0 is the average density of condensed bosons. This marked dependence of the Meissner-Ochsenfeld effect on the boundary conditions reflects the fact that the ideal Bose gas without any interactions whatsoever is a highly idealized system. However, it will be shown in Sec. 4 that this arbitrariness can be removed by including, in a self-consistent way, the Coulomb interaction between the bosons. In this case, the London equation (1.14) is restored.

What the relation of the present model to the future theory of superconductivity will be, is hard to foresee. It seems, however, unlikely that there should be no close relation between the two, especially in view of the fact that the Bose-Einstein condensation of a boson gas—which gives the transition to the superconducting state of the model—is a very singular and unique phenomenon which is responsible also for the other spectacular low-temperature phenomenon: superfluidity.[10] It seems therefore reasonable to expect superconductivity in metals to be due to the occurrence of charge-carrying bosons in the metal. The author has recently suggested[11] that these "bosons" might be resonant two-electron states. A crude theory of chemical equilibrium between these "bosons" and free electrons then gives results which agree qualitatively (although not quantitatively) with the observed behavior. No detailed discussion of this suggestion, however, will be given in this paper.

2. INDUCED CURRENT DENSITY

In order to calculate the current density induced by a weak inhomogeneous magnetic field in a charged ideal gas, we use perturbation theory on the distribution function as developed earlier.[3,12] As we are considering inhomogeneous fields, perturbation theory on the magnetic field is here permitted, in contrast to the case of homogeneous fields, where arbitrarily small fields produce qualitative changes in the wave functions of the electrons, as soon as the volume is big enough. Also, perturbation theory on the distribution function does not suffer from the limitations of perturbation theory on the energy levels in the case of complete or near degeneracies.[12,13]

A function $F_0(\Lambda_0+\epsilon\Lambda_1)$ of an operator $\Lambda\equiv\Lambda_0+\epsilon\Lambda_1$ is expanded in powers of ϵ in the representation in which Λ_0 is diagonal

$$\Lambda_0|n\rangle=\lambda_n|n\rangle, \tag{2.1}$$

by

$$\langle n|F_0(\Lambda_0+\epsilon\Lambda_1)|m\rangle=\delta_{nm}F_0(\lambda_n)+\epsilon\langle n|\Lambda_1|m\rangle F_1(\lambda_n,\lambda_m)+O(\epsilon^2), \tag{2.2}$$

where

$$F_1(\lambda_n,\lambda_m)=\frac{F_0(\lambda_n)-F_0(\lambda_m)}{\lambda_n-\lambda_m}. \tag{2.3}$$

We take Λ_0 to be the Hamiltonian of a free particle

$$\Lambda_0\equiv H_0=(1/2m)\mathbf{p}^2, \tag{2.4}$$

and $\epsilon\Lambda_1$ to be the perturbation due to a magnetic field

$$\epsilon\Lambda_1\equiv H'=-\frac{e}{2mc}\{\mathbf{p}\cdot\mathbf{A}(\mathbf{x})+\mathbf{A}(\mathbf{x})\cdot\mathbf{p}\}+O(\mathbf{A}^2), \tag{2.5}$$

where $\mathbf{A}(\mathbf{x})$ is the vector potential of the magnetic field in an arbitrary gauge. $F_0(H)(H=H_0+H')$ is the distribution function, i.e., for a Bose gas

$$F_0(H)=[e^{\alpha(H-\mu)}-1]^{-1} \tag{2.6}$$

($\alpha=1/kT$, $\mu<0$ chemical potential).

The average current density in thermal equilibrium is then, up to terms linear in the magnetic field[14]

$$\mathbf{i}(\mathbf{x})=-\frac{e^2}{m^2c}\frac{1}{G}\sum_{\mathbf{p},\mathbf{q}}(\mathbf{p}\cdot\mathbf{A}(\mathbf{q}))\mathbf{p}\times\exp(-i\mathbf{q}\cdot\mathbf{x})\cdot F_1\left[\frac{(\mathbf{p}+\mathbf{q}/2)^2}{2m},\frac{(\mathbf{p}-\mathbf{q}/2)^2}{2m}\right]-\frac{e^2}{mc}\frac{1}{G}\mathbf{A}(\mathbf{x})\sum_{\mathbf{p}}F_0\left(\frac{\mathbf{p}^2}{2m}\right). \tag{2.7}$$

Here $\mathbf{A}(\mathbf{q})$ is the Fourier transform of $\mathbf{A}(\mathbf{x})$:

$$\mathbf{A}(\mathbf{x})=\sum_{\mathbf{q}}\mathbf{A}(\mathbf{q})\exp(-i\mathbf{q}\cdot\mathbf{x}). \tag{2.8}$$

Simplifying (2.7) using

$$\sum_{\mathbf{p}}F_0\left(\frac{\mathbf{p}^2}{2m}\right)=N, \tag{2.9}$$

we get

$$i_\mu(\mathbf{x})=-\frac{e^2}{mc}\frac{N}{G}A_\mu(\mathbf{x})-\frac{e^2}{mc}\frac{1}{G}\sum_{\nu=1}^{3}\sum_{\mathbf{q}}I_{\mu\nu}(\mathbf{q})A_\nu(\mathbf{q})\exp(-i\mathbf{q}\cdot\mathbf{x}), \tag{2.10}$$

where

$$I_{\mu\nu}(\mathbf{q})=\sum_{\mathbf{p}}\frac{p_\mu p_\nu}{(\mathbf{p}\cdot\mathbf{q})}\left[F_0\left(\frac{(\mathbf{p}+\mathbf{q}/2)^2}{2m}\right)-F_0\left(\frac{(\mathbf{p}-\mathbf{q}/2)^2}{2m}\right)\right], \tag{2.11}$$

and where use has been made of (2.3).

In order to evaluate $I_{\mu\nu}(\mathbf{q})$ we exploit the fact that it depends only on the vector $\mathbf{q}$, so that it must have the

[10] F. London, Phys. Rev. **54**, 947 (1938); S. T. Butler and M. H. Friedman, Phys. Rev. **98**, 287 (1955); J. M. Blatt and S. T. Butler, Phys. Rev. **96**, 1149 (1954).

[11] M. R. Schafroth, Phys. Rev. **96**, 1442 (1954); see also C. J. Gorter, *Progress in Low Temperature Physics* (Interscience Publishers, Inc., New York, 1955), Vol. I.

[12] M. J. Buckingham and M. R. Schafroth, Proc. Phys. Soc. (London) **A67**, 828 (1954).

[13] R. Peierls, Z. Physik **80**, 763 (1933).

[14] We put $\hbar=1$ throughout this section for convenience of notation.

form

$$I_{\mu\nu}(\mathbf{q})=a(q^2)q_\mu q_\nu+b(q^2)q^2\delta_{\mu\nu}. \tag{2.12}$$

Then:

$$\sum_\mu I_{\mu\mu}=(a+3b)q^2\equiv I_0,$$
$$\sum_{\mu\nu} I_{\mu\nu}q_\mu q_\nu=(a+b)q^4\equiv I_1, \tag{2.13}$$

where now

$$I_0=\sum_{\mathbf{p}}\frac{\mathbf{p}^2}{(\mathbf{p}\cdot\mathbf{q})}\left\{F_0\left[\frac{(\mathbf{p}+\mathbf{q}/2)^2}{2m}\right]-F_0\left[\frac{(\mathbf{p}-\mathbf{q}/2)^2}{2m}\right]\right\},$$
$$I_1=\sum_{\mathbf{p}}(\mathbf{p}\cdot\mathbf{q})\left\{F_0\left[\frac{(\mathbf{p}+\mathbf{q}/2)^2}{2m}\right]-F_0\left[\frac{(\mathbf{p}-\mathbf{q}/2)^2}{2m}\right]\right\}. \tag{2.14}$$

I_1 gives by straightforward evaluation

$$I_1=\sum_{\mathbf{p}}\left[\left(\mathbf{p}-\frac{\mathbf{q}}{2}\right)\mathbf{q}-\left(\mathbf{p}+\frac{\mathbf{q}}{2}\right)\mathbf{q}\right]F_0\left(\frac{\mathbf{p}^2}{2m}\right)$$
$$=-q^2\sum_{\mathbf{p}}F_0\left(\frac{\mathbf{p}^2}{2m}\right)$$

or

$$I_1=-q^2N. \tag{2.15}$$

Using this, one finds from (2.12) and (2.13)

$$I_{\mu\nu}(\mathbf{q})=\tfrac{1}{2}(I_0+N)\left(\delta_{\mu\nu}-\frac{q_\mu q_\nu}{q^2}\right)-\frac{N}{q^2}q_\mu q_\nu,$$

and upon inserting into (2.10)

$$i_\mu(\mathbf{x})=-\frac{e^2}{mc}\frac{1}{G}\sum_{\nu=1}^{3}\sum_{\mathbf{q}}A_\nu(\mathbf{q})\exp(-i\mathbf{q}\cdot\mathbf{x})$$
$$\times\tfrac{1}{2}(I_0+3N)\left(\delta_{\mu\nu}-\frac{q_\mu q_\nu}{q^2}\right), \tag{2.16}$$

which exhibits gauge invariance explicitly. Comparison with (1.10) and (1.11) shows

$$K(q^2)=\frac{e^2}{2mc}\frac{1}{G}\frac{I_0+3N}{q^2}. \tag{2.17}$$

We now proceed to evaluate I_0:

(1) For a noncondensed Bose gas (as well as for a Fermi gas) the distribution function is regular at $\mathbf{p}^2=0$ so that for small $|\mathbf{q}|$ we may expand the summand in (2.14) in powers of $\mathbf{q}$. We thus write

$$F_0\left(\frac{(\mathbf{p}+\mathbf{q}/2)^2}{2m}\right)-F_0\left(\frac{(\mathbf{p}-\mathbf{q}/2)^2}{2m}\right)$$
$$=\frac{\mathbf{p}\cdot\mathbf{q}}{m}F_0'\left(\frac{\mathbf{p}^2}{2m}\right)+\frac{\mathbf{p}\cdot\mathbf{q}}{m}\frac{\mathbf{q}^2}{8m}F_0''\left(\frac{\mathbf{p}^2}{2m}\right)$$
$$+\tfrac{1}{3}\left(\frac{\mathbf{p}\cdot\mathbf{q}}{2m}\right)^3F_0'''\left(\frac{\mathbf{p}^2}{2m}\right)+O(\mathbf{q}^6), \tag{2.18}$$

where the dash indicates differentiation with respect to the energy $E=\mathbf{p}^2/2m$. The first term gives

$$I_{00}=2\sum_{\mathbf{p}}\frac{\mathbf{p}^2}{2m}F_0'\left(\frac{\mathbf{p}^2}{2m}\right),$$

which in turn, replacing sums by integrals, yields

$$I_{00}=\frac{G}{(2\pi)^3}2\int d^3\mathbf{p}\frac{\mathbf{p}^2}{2m}F_0'\left(\frac{\mathbf{p}^2}{2m}\right)$$
$$=\frac{G}{4\pi^2}2(2m)^{\frac{1}{2}}\int_0^\infty dEE^{\frac{1}{2}}F_0'(E)$$

and by integration by parts

$$I_{00}=-\frac{G}{4\pi^2}3(2m)^{\frac{1}{2}}\int_0^\infty dEE^{\frac{1}{2}}F_0(E)=-3\sum_{\mathbf{p}}F_0\left(\frac{\mathbf{p}^2}{2m}\right),$$

or

$$I_{00}=-3N. \tag{2.19}$$

The remaining terms in (2.18) give similarly

$$I_{02}=\frac{\mathbf{q}^2}{12m}\sum_{\mathbf{p}}\left(\frac{\mathbf{p}^2}{2m}\right)^{-1}F_0\left(\frac{\mathbf{p}^2}{2m}\right)=\frac{q^2}{12m}N\langle E^{-1}\rangle_{\text{kin}}, \tag{2.20}$$

so that

$$K(q^2)=\frac{e^2}{24m^2c}\frac{N\langle E^{-1}\rangle_{\text{kin}}}{G}+O(q^2), \tag{2.21}$$

i.e., we get the usual diamagnetism (1.5) of a gas of charged particles, but no Meissner effect.

(2) For a condensed Bose-gas, however, the distribution function is no longer regular at $\mathbf{p}^2=0$ and, therefore, the above procedure does not apply. The ground state is highly occupied (by N_0 particles) and has to be treated separately[6]; for the other states one can use the usual Bose distribution with $\mu=0$; replacement of sums by integrals is justified.

We write

$$I_0=\sum_{\mathbf{p}}\left[\frac{(\mathbf{p}-\mathbf{q}/2)^2}{(\mathbf{p}-\mathbf{q}/2)\cdot\mathbf{q}}-\frac{(\mathbf{p}+\mathbf{q}/2)^2}{(\mathbf{p}+\mathbf{q}/2)\cdot\mathbf{q}}\right]F_0\left(\frac{\mathbf{p}^2}{2m}\right). \tag{2.22}$$

The noncondensed particles contribute

$$I_{0n}=\frac{G}{(2\pi)^3}\int d^3\mathbf{p}\left[\frac{(\mathbf{p}-\mathbf{q}/2)^2}{(\mathbf{p}-\mathbf{q}/2)\cdot\mathbf{q}}-\frac{(\mathbf{p}+\mathbf{q}/2)^2}{(\mathbf{p}+\mathbf{q}/2)\mathbf{q}}\right]$$
$$\times\left[\exp\left(\alpha\frac{\mathbf{p}^2}{2m}\right)-1\right]^{-1}. \tag{2.23}$$

Care has to be taken here because, in transforming I_0 from the original form (2.14) to (2.22), singularities in the two terms under the sum which compensated in (2.14) have been separated, so that (2.23) contains singularities. A prescription for integrating over these

singularities must therefore be given which ensures the equivalence of (2.14) and (2.23). It can be seen that it is a correct prescription in this sense to take the principal value in the integration over the angles in (2.23).

Performing the integration over the angles in (2.23) then yields

$$I_{0n}=4\pi\frac{G}{(2\pi)^3}\int_0^\infty dp\left[\frac{p}{q}\left(p^2-\frac{q^2}{4}\right)\times\log\left|\frac{p-q/2}{p+q/2}\right|-2p^2\right]\left[\exp\left(\alpha\frac{p^2}{2m}\right)-1\right]^{-1}$$

with $q=|\mathbf{q}|$.

The number of noncondensed bosons is

$$N_n\equiv N-N_s=\sum_{\mathbf{p}\neq 0}\left[\exp\left(\alpha\frac{\mathbf{p}^2}{2m}\right)-1\right]^{-1}=4\pi\frac{G}{(2\pi)^3}\int_0^\infty dp\,p^2\left[\exp\left(\alpha\frac{p^2}{2m}\right)-1\right]^{-1}$$

so that

$$I_{0n}+3N_n=4\pi\frac{G}{(2\pi)^3}\int_0^\infty dp\left[\frac{p}{q}\left(p^2-\frac{q^2}{4}\right)\times\log\left|\frac{p-q/2}{p+q/2}\right|+p^2\right]\times\left[\exp\left(\alpha\frac{p^2}{2m}\right)-1\right]^{-1}. \quad (2.24)$$

The contribution from the condensed particles is, from (2.22), for $p=0$:

$$I_{0s}=-N_s,$$

and

$$I_{0s}+3N_s=2N_s. \quad (2.25)$$

Thus finally, inserting (2.24) and (2.25) into (2.17) we get

$$K(q^2)=K_s(q^2)+K_n(q^2),$$

$$K_s(q^2)=\frac{e^2}{mc}\frac{N_s}{G}\frac{1}{q^2},$$

$$K_n(q^2)=\frac{e^2}{mc}\frac{1}{4\pi^3}\frac{1}{q^2}\int_0^\infty dp\left[\frac{p}{q}\left(p^2-\frac{q^2}{4}\right)\times\log\left|\frac{p-q/2}{p+q/2}\right|+p^2\right]\left[\exp\left(\alpha\frac{p^2}{2m}\right)-1\right]^{-1}. \quad (2.26)$$

It is seen that the integral in the second term vanishes for $q\to 0$, so that $K_s(q^2)$ cannot be compensated by $K_n(q^2)$. $K_s(q^2)$, the contribution from the condensed bosons, has exactly the form (1.13) of London's equation. $K_n(q^2)$ stems from the uncondensed particles and vanishes at absolute zero so that for $T=0$ London's equation holds exactly. At finite temperatures, $K_n(q^2)$ can be written[15]

$$K_n(q^2)=\frac{e^2}{2mc}\frac{1}{4\pi^3}\frac{(2mkT)^{\frac{1}{2}}}{\hbar^3q^2}k\left(\frac{\hbar q}{2(2mkT)^{\frac{1}{2}}}\right), \quad (2.27)$$

where

$$k(z)\equiv\int_0^\infty dx\left[\frac{x}{2z}-(x^2-z^2)\log\left|\frac{x-z}{x+z}\right|+x^2\right]\frac{1}{e^{x^2}-1}. \quad (2.28)$$

It is shown in Appendix I that the expansion of $k(z)$ for small $|z|$ is

$$k(z)=\frac{\pi^2}{4}z+\frac{\sqrt{\pi}}{3}\zeta(\tfrac{1}{2})z^2+O(z^4). \quad (2.29)$$

The second and higher terms in this expansion give simply a normal diamagnetism (1.12), whereas the first term still yields a pole of $K_n(q^2)$ at $q=0$:

$$K_n(q^2)=\frac{1}{32}\frac{e^2}{mc}\frac{2mkT}{\hbar^2q}+O(1). \quad (2.30)$$

Thus finally, from (2.26) and (2.30):

$$K(q^2)=\frac{e^2}{mc}\left[\frac{N_s}{G}\frac{1}{q^2}+\frac{1}{32}\frac{2mkT}{\hbar^2q}+O(1)\right]. \quad (2.31)$$

3. DISCUSSION OF THE KERNEL $K(q^2)$

As pointed out earlier, the form (2.31) of $K(q^2)$ shows that the relation between field and induced current does not take the simple London form (1.13), but it is rather an integral relationship. It will be shown, however, that $K_n(q^2)$ is in general negligible as compared to $K_s(q^2)$, i.e., that London's equation holds with a very good accuracy for the condensed Bose gas.

One has to judge the relative importance of the two terms in (2.31) for values of $q\sim d^{-1}$, where d is the penetration depth. We take for d the value given by the London term $K_s(q^2)$ alone, which is correct as long as $K_n(q^2)$ is indeed negligible, i.e., we write[9]

$$d=(mc^2/4\pi n_se^2)^{\frac{1}{2}}. \quad (3.1)$$

$K_n(q^2)$ is then negligible compared to $K_s(q^2)$ as long as

$$n_s\gg\frac{\pi}{256}\frac{e^2}{mc^2}\left(\frac{2mkT}{\hbar^2}\right)^2. \quad (3.2)$$

Using (1.4) this yields:

$$1-(T/T_c)^{\frac{3}{2}}\gg\frac{\pi^3}{16[\zeta(\frac{3}{2})]^{4/3}}\frac{e^2}{mc^2}n^{\frac{1}{3}}. \quad (3.3)$$

This shows that the deviation of the magnetic behavior of a Bose gas from the pure London equation is negligible as long as the mean distance between bosons

[15] We reinsert here for convenience the constant $\hbar$ which was put $=1$ before.

$(n^{-\frac{1}{3}})$ is large compared to the "classical boson radius" e^2/mc^2. Fitting, e.g., n through (1.1) to a transition temperature of a few degrees Kelvin and taking the mass and charge of the bosons to be of the order of the mass and charge of electrons, the right-hand side of (3.3) is of the order 10^{-8}.

We therefore conclude that for all practical purposes the condensed Bose gas obeys London's equation

$$-\lambda c\,\mathrm{curl}\mathbf{i}=\mathbf{H}, \tag{3.4}$$

with

$$\lambda^{-1}=(e^2/m)n_s, \tag{3.5}$$

n_s being the number of condensed bosons per unit volume.

The penetration depth d is given by (3.1), and its temperature dependence is

$$d=d_0[1-(T/T_c)^4]^{-\frac{1}{2}}, \tag{3.6}$$

with

$$d_0=(mc^2/4\pi e^2 n)^{\frac{1}{2}} \tag{3.7}$$

n being the total boson density.

In the temperature region in which (3.3) is violated the penetration depth becomes larger than the value (3.6). It can be shown that the definition of the penetration depth given, e.g., by Pippard[16] is equivalent to the following definition in terms of $f(q)=q^2K(q^2)$:

$$d^{-1}=\frac{2}{\pi}\int_0^\infty dq\frac{(4\pi/c)f(0)}{q^2+(4\pi/c)f(q)}. \tag{3.8}$$

This shows that the $1/q$-term as well as a normal diamagnetism [(1.12) with $\chi'<0$] tend to increase the penetration depth. One finds from (3.8) that, as $T\to T_c$, d actually becomes infinite like

$$\{(1-T/T_c)\log(1-T/T_c)\}^{-1},$$

instead of the behavior $(1-T/T_c)^{-\frac{1}{2}}$ which arises out of (3.6).

4. INFLUENCE OF THE BOUNDARY CONDITION

The complete analysis carried out in Sec. 2 for the cubical volume with periodical boundary conditions becomes very cumbersome if applied to other volumes. We are therefore going to use a simplified approach to study the dependence on the boundary conditions. A convenient volume to use is a cube of volume L^3 with walls at $x=0$ and $x=L$; in the y- and z-directions we shall apply periodical boundary conditions. This volume, though still artificial, allows a study of the effect of walls in a simple way without introducing spurious complications due to corners, etc. In addition, we restrict ourselves to an inhomogeneous magnetic field depending on x alone; from $\mathrm{div}B=0$ it follows that, if we exclude a homogeneous component parallel to x, $\mathbf{B}$ must be parallel to the wall; we take $\mathbf{B}$ to lie in the y-direction. $\mathbf{B}$ can then be described by a vector potential $\mathbf{A}$ in a gauge where

$$\mathrm{div}\mathbf{A}=0, \tag{4.1}$$

and on the surface

$$A_{\mathrm{normal}}=0. \tag{4.2}$$

This entails

$$A_x=A_y=0;\quad A_z=A(x). \tag{4.3}$$

If we now neglect the contributions of the noncondensed bosons, our problem reduces to calculating the first-order perturbation induced by a vector potential (4.3) in the current density carried by the ground state.

The states of the system are given by

$$\psi_{k_1k_2n}(\mathbf{x})=(1/L)e^{i(k_1y+k_2z)}\varphi_n(x), \tag{4.4}$$

where $k_i=(2\pi/L)s_i$, s_i being integers, and $\varphi_n(x)$ are normalized one-dimensional free-particle wave functions in the interval $0\le x\le L$. They are determined by boundary conditions at $x=0$ and $x=L$. As it is just the effect of these boundary conditions that we are interested in, we are going to leave them open for the time being.

The ground-state wave function is

$$\psi_{000}(\mathbf{x})=(1/L)\varphi_0(x), \tag{4.5}$$

and the local density of condensed bosons is

$$\begin{aligned} n_s(\mathbf{x})&=Ln_s^0|\varphi_0(x)|^2,\\ n_s^0&=L^{-3}\int d^3\mathbf{x}n_s(\mathbf{x}). \end{aligned} \tag{4.6}$$

The perturbation operation in the Hamiltonian due to the field (4.3) is

$$H^1=-\frac{eh}{mc}(1/i)A(x)\frac{\partial}{\partial z}. \tag{4.7}$$

The matrix element which determines the perturbation of the ground-state wave function is

$$(k_1k_2n|H^1|000), \tag{4.8}$$

and this is *identically zero*, irrespective of the form of $\varphi_0(x)$, i.e., for all kinds of boundary conditions at the walls.

The current density

$$\mathbf{i}(\mathbf{x})=n_s^0\left\{\frac{eh}{2mi}\left(\frac{\partial\psi^*}{\partial\mathbf{x}}\psi-\psi^*\frac{\partial\psi}{\partial\mathbf{x}}\right)-\frac{e^2}{mc}\mathbf{A}(\mathbf{x})|\psi(\mathbf{x})|^2\right\}, \tag{4.9}$$

therefore becomes, using (4.6)

$$\mathbf{i}(\mathbf{x})=-(e^2/mc)\mathbf{A}(\mathbf{x})n_s(\mathbf{x}), \tag{4.10}$$

which obeys (1.14′) and (1.15′).

This is exactly the way in which London[9] predicted superconductivity to arise: A long-range order (due here to the nature of the condensed state) prevents the wave function of the superconducting particles from adapting itself to the magnetic field. It must, however, be noted

[16] A. B. Pippard, Proc. Roy. Soc. (London) **A216**, 547 (1953).

that such an argument is valid only if the expression (4.9) for the current density holds, which will in general not be true if the long-range order is due to some interactions. It has, for example, been shown very generally for weak-coupling interactions that any such tendency of the wave function to "stiffen" in that way will be countered by extra terms in the current density which nullify the effect.[3,17] In the Bose gas, however, the long-range order is a purely statistical effect, so that (4.9) holds; and therefore London's argument goes through.

We now specify boundary conditions:

(a) For perfectly reflecting walls,

$$\psi(0)=\psi(L)=0, \tag{4.11}$$

we have

$$\varphi_0(x)=(2/L)^{\frac{1}{2}}\sin(\pi x/L), \tag{4.12}$$

and

$$n_s(x)=n_0^s(2/L)\sin^2(\pi x/L). \tag{4.13}$$

Inserted into (4.10), this gives a law of penetration which depends on the size of the container, in disagreement with experimental facts.

(b) The boundary condition

$$\left(\frac{\partial\psi}{\partial x}\right)_{x=0}=\left(\frac{\partial\psi}{\partial x}\right)_{x=L}=0, \tag{4.14}$$

or the periodicity condition

$$\begin{aligned}&\psi(0)=\psi(L),\\ &\left(\frac{\partial\psi}{\partial x}\right)_{x=0}=\left(\frac{\partial\psi}{\partial x}\right)_{x=L},\end{aligned} \tag{4.15}$$

on the other hand yield

$$\varphi_0(x)=L^{-\frac{1}{2}}, \tag{4.16}$$

$$n_s(x)=n_s^0, \tag{4.17}$$

i.e., the London equation (1.14) holds.

As mentioned in the introduction, this strong dependence of the results on the boundary conditions is due to the fact that the ideal Bose gas with no interaction whatsoever between particles is physically quite unrealistic. It will, however, still provide a useful model of a superconductor if one chooses artificially the boundary condition (4.14) which yields a uniform local density $n_s(\mathbf{x})$ of condensed particles for any shape of the volume.

By a qualitative argument we shall now show that this is, indeed, the proper procedure if one takes into account the Coulomb repulsion between particles.

In order to include the Coulomb interaction between bosons in a self-consistent way, we shall treat the particles as moving in a potential $V(x)$. $V(x)$ is related by Poisson's equation to the local charge density of the bosons; a uniform background charge has to be subtracted to make the whole system electrically neutral. The self-consistent equations for the wave function φ of the condensed state and the potential $V(x)$ are

$$-\frac{\hbar^2}{2m}\frac{d^2\varphi}{dx^2}+[V(x)-E]\varphi(x)=0, \tag{4.18}$$

$$\frac{d^2V}{dx^2}+4\pi e^2 n_s^0(L|\varphi(x)|^2-1)=0. \tag{4.19}$$

In (4.19), $\rho_s=e^2n_s^0L|\varphi|^2$ is the charge density of the condensed bosons, whereas $\rho_n=-e^2n_s$ is the charge density of the uniform background and of the noncondensed particles.[18] E is the energy shift of the state $\varphi(x)$ due to the Coulomb interaction; it is linked with the normalization condition

$$\int_0^L dx|\varphi(x)|^2=1, \tag{4.20}$$

since the system (4.18), (4.19) is nonlinear.

The system (4.18), (4.19) has to be solved subject to the boundary conditions

$$\begin{aligned}&\varphi(0)=\varphi(L)=0,\\ &V(0)=V(L)=0,\end{aligned} \tag{4.21}$$

and the normalization condition (4.20). We are only interested in the ground-state solution, i.e., the one for which $\varphi(x)$ has no nodes. This excludes an oscillatory behavior of $\varphi(x)$ and therefore, through (4.19), of $V(x)$. Furthermore, $\varphi(x)$ cannot be appreciably different from its average value $L^{-\frac{1}{2}}$ over large regions (i.e., regions comparable to L), since this would lead to a building-up of $V(x)$ through (4.19), which in turn contradicts (4.18). We therefore expect $\varphi(x)$ to be practically equal to $L^{-\frac{1}{2}}$ throughout most of the range $0<x<L$, with a small region near the boundaries where it drops to its boundary value zero. We are now going to estimate this region near the boundary and show that the picture of the solution we thus get is indeed consistent.

We therefore put

$$\begin{aligned}&\varphi(x)=L^{-\frac{1}{2}}, \quad \lambda<x<L-\lambda,\\ &V(x)=V_0.\end{aligned} \tag{4.22}$$

(4.22) is consistent with (4.18), (4.19) if we put $E=V_0$; V_0 is, therefore, just the energy shift of the ground-state due to the Coulomb repulsion.

The normalization condition (4.21) now requires that the average value of $|\varphi(x)|^2$ in the boundary regions $0<x<\lambda$ and $L-\lambda<x<L$ be equal to L:

$$\int_0^\lambda dx|\varphi(x)|^2=\lambda/L. \tag{4.23}$$

[17] M. R. Schafroth, Nuovo cimento **9**, 291 (1952).

[18] Strictly, the charge density of the noncondensed bosons is not uniform if $V(x)$ is not constant. However, ρ_n is much less sensitive to $V(x)$ than the density ρ_s of condensed bosons, and since, as we shall show, the resulting $V(x)$ tends to be smoothed out, the approximation $\rho_n=$ constant is sufficient for our purpose.

This means that $\varphi(x)$ must rise from its value $\varphi(0)=0$ to a peak and then drop to the value $\varphi(\lambda)=L^{-\frac{1}{2}}$, and by symmetry, similarly near $x=L$. (4.19) then shows that $V(x)$ rises from $V(0)=0$ to a value

$$V_0\equiv V(\lambda)\approx n_s^0\cdot e^2\cdot\lambda^2. \tag{4.24}$$

(4.18) now requires that

$$\hbar^2/2m\lambda^2\approx E-V(0)=V(\lambda)-V(0)\approx n_s^0e^2\lambda^2. \tag{4.25}$$

Consistency is thus achieved for a λ of the order of magnitude

$$\lambda^4\approx\hbar^2/me^2n_s^0. \tag{4.26}$$

Using (3.1) to relate n_s^0 to the penetration depth d we find

$$\lambda/d\approx\left(\frac{h/mc}{d}\right)^{\frac{1}{2}}. \tag{4.27}$$

For bosons of a mass comparable to or larger than the electron mass the right-hand side is always very small, so that the layer λ can always be neglected for calculations about superconducting phenomena. For our purposes, therefore, the Bose gas including Coulomb interaction behaves like an ideal Bose gas with the artificial boundary condition

$$(\partial\psi/\partial n)_{\text{wall}}=0. \tag{4.28}$$

(4.28) gives the correct ground state $\psi(x)=G^{-\frac{1}{2}}$ for all volumes; the contributions due to the higher states, i.e., the contributions of the noncondensed bosons, are not affected by the choice of the boundary condition anyway, apart from pure surface effects.

The strong flattening of the ground-state wave function due to the Coulomb repulsion between particles may be surprising at first sight. It should be noted that this is again due to the high occupation of the ground state by the condensed particles: All of them contribute simultaneously to the potential $V(x)$ acting on any single one, and this cumulative effect is responsible for the high sensitivity of $V(x)$ to $\varphi(x)$ [through the factor n_s^0L in front of $|\varphi(x)|^2$ in (4.19)].

5. THE CRITICAL FIELD

Let us now consider a Bose gas in a cylindrical container in an "acting" homogeneous magnetic field H' parallel to the axis of the container. The grand canonical partition function $e^{-\alpha\Omega}$ can be worked out in a standard way[19] to give

$$\Omega=kT\frac{eH'}{2hc}\frac{G}{2\pi^2}\int_{-\infty}^{+\infty}d\kappa\sum_{\nu=0}^{\infty}\times\log\left\{1-\exp\left[-\alpha\left(\mu_0H'(2\nu+1)+\frac{\hbar^2\kappa^2}{2m}-\mu\right)\right]\right\}, \tag{5.1}$$

where G is the volume of the container, $\mu_0=eh/2mc$ Bohr's magneton, $\mu=$ the chemical potential, $\alpha=1/kT$.

The number of particles is given by

$$N=-\frac{\partial\Omega}{\partial\mu}=\frac{eH'}{2hc}\frac{G}{2\pi^2}\int_{-\infty}^{+\infty}d\kappa\sum_{\nu=0}^{\infty}\left\{\exp\left[\alpha\left(\mu_0H'(2\nu+1)+\frac{\hbar^2\kappa^2}{2m}-\mu\right)\right]-1\right\}^{-1}. \tag{5.2}$$

The free energy is

$$F=\Omega+\mu N. \tag{5.3}$$

The chemical potential μ is here no longer restricted to negative values as in the absence of a magnetic field, but can take values up to

$$\mu_{\max}=+\mu_0H'. \tag{5.4}$$

(5.1) and (5.2) can be brought into a more tractable form by expanding the integrands in powers of $\exp\{-\alpha[\mu_0H'(2\nu+1)+(\hbar^2\kappa^2/2m)-\mu]\}$ and interchanging summation and integration; this yields

$$\omega\equiv\frac{\Omega}{G}=-kTn_0\frac{1}{\zeta(\frac{3}{2})}\sum_{\nu=1}^{\infty}\frac{e^{\alpha\mu\nu}}{\nu^{3/2}}\frac{\alpha\mu_0H'\nu}{\sinh(\alpha\mu_0H'\nu)}, \tag{5.5}$$

$$n\equiv\frac{N}{G}=n_0\frac{1}{\zeta(\frac{3}{2})}\sum_{\nu=1}^{\infty}\frac{e^{\alpha\mu\nu}}{\nu^{\frac{1}{2}}}\frac{\alpha\mu_0H'\nu}{\sinh(\alpha\mu_0H'\nu)}. \tag{5.6}$$

Here,

$$n_0=\zeta(\tfrac{3}{2})\left(\frac{2mkT}{4\pi\hbar^2}\right)^{\frac{3}{2}} \tag{5.7}$$

is the maximum density of particles that can be accommodated outside the ground state in the Bose gas at temperature T without magnetic field.

On putting $H'=0$, (5.5) and (5.6) reduce to well-known expressions for the free Bose gas.

However, even arbitrarily small values of the magnetic field H' introduce qualitative changes. However small H', the sum in (5.6) can take arbitrarily large values, so that any density of bosons can be accommodated outside the ground state at all finite temperatures: *The Bose gas does not condense at any finite temperature if it is in a fixed homogeneous magnetic field.*[20]

An important question at this point is the precise nature of the "acting" field H'. The Lorentz relation

$$H'=H+(4\pi/3)M, \tag{5.8}$$

[19] W. Pauli, Solvay-Report 1930 (Gauthier-Villars, Paris, 1932), pp. 183–190 and H. A. Bethe and A. Sommerfeld, *Handbuch der Physik* (Verlag Julius Springer, Berlin, 1933), Vol. 24, Part 2, pp. 477 ff.

[20] It should be kept in mind here that (5.1), (5.2), (5.5), (5.6) are derived under the assumption that the volume G tends to infinity at fixed H', i.e., that

$$(e/c)H'\gg(2mkT)^{\frac{1}{2}}/R\ (A),$$

where R is the radius of the cylindrical container. The above statement therefore holds only for fields fulfilling (A), and "arbitrarily small fields" strictly means "fields of the order $(2mkT)^{\frac{1}{2}}/R$."

where H is the applied field, $M=-\partial\Omega(H')/\partial H'$ the magnetization, cannot be applied here. In fact, an estimate (see Appendix II) shows that for all values of H' which need be considered, the diameter of the bosons' orbits are much larger than the mean distance between particles, provided that the right-hand side of (3.3) is small, i.e., provided that

$$\eta\equiv(e^2/mc^2)n^{\frac{1}{3}}\ll 1. \tag{5.9}$$

The field produced by any one such large orbit is negligible compared to the average field, and, therefore, the "acting" field H' has to be identified with the average microscopic field B.[21]

$$H'=B=H+4\pi M \tag{5.10}$$

It will be shown at the end of this section that, apart from negligible corrections, for $T<T_c(1-\eta)$:

$$M(H')=-n_s\mu_0\frac{H'}{|H'|}$$

$$=-n\mu_0\frac{H'}{|H'|}\left[1-\left(\frac{T}{T_c}\right)^{\frac{3}{2}}\right], \tag{5.11}$$

where μ_0 is the Bohr magneton. In a magnetic field B, therefore, the Bose gas exhibits a permanent diamagnetic moment, it is a "dia-ferromagnet." No condensed phase exists.

From (5.10) and (5.11), it can be seen that the $B(H)$-curve for the Bose-gas is given by

$$B(H)=\begin{cases}0 & H<H_0\\ H-H_0 & H>H_0\end{cases}, \tag{5.12}$$

with

$$H_0=4\pi n\mu_0\left[1-(T/T_c)^{\frac{3}{2}}\right]. \tag{5.13}$$

This means that for $H>H_0$ a field H applied to the surface of the cylinder penetrates the Bose gas as a *homogeneous* field, producing a magnetization M and an induction B related by (5.10) and (5.11). For $H<H_0$, however, no *homogeneous* field can penetrate the Bose gas; indeed, assuming a homogeneous $H<H_0$ leads to a contradiction in (5.10) and (5.11). If one applies to the surface of the cylinder an $H<H_0$ the bosons condense, expel the field except for a thin surface layer—inhomogeneous fields are, of course, outside the scope of the approach in this section—and the superconducting state is established. $H_0(T)$ (5.12) is, therefore, the critical field. Using (3.6) and (3.7), $H_0(T)$ can be related to the penetration depth d:

$$H_c=\hbar c/2ed^2. \tag{5.14}$$

[21] From (5.10) one can deduce the following relation between the susceptibility χ and the quantity χ' of (1.5):

$$\chi=\chi'/(1-4\pi\chi').$$

This shows that on approaching condensation, where $\chi'\to-\infty$, $\chi\to-1/4\pi$ so that the permeability tends to zero, in accordance with (5.12). See also H. Fröhlich, Nature **168**, 280 (1951).

The $B(H)$ law (5.12) of the Bose gas differs from the one usually assumed for superconductors,[22] and therefore the thermodynamical relations have to be reinvestigated.

Calling $\Phi(T)$ the free energy of the Bose gas at vanishing field, the thermodynamical potential with variables T and H, $G(H,T)$,[23] is given by

$$G(H,T)=\Phi(T)-\frac{1}{4\pi}\int_0^H BdH$$

and hence

$$\begin{aligned}&G_s(H,T)=\Phi(T);\quad (H<H_0),\\ &G_n(H,T)=\Phi(T)-(1/8\pi)(H-H_0)^2;\quad (H>H_0).\end{aligned} \tag{5.15}$$

From this, one finds for the latent heat of the transition in the field

$$Q=-T\frac{\partial}{\partial T}(G_n-G_s)_{H=H_0}=\frac{1}{8\pi}T\left.\frac{\partial(H-H_0)^2}{\partial T}\right|_{H=H_0}. \tag{5.16}$$

The jump in the specific heat is, similarly

$$C_n-C_s=-T\frac{\partial^2}{\partial T^2}(G_n-G_s)_{H=H_0}$$

$$=\frac{T}{8\pi}\frac{\partial^2}{\partial T^2}(H-H_0)^2\big|_{H=H_0}, \tag{5.17}$$

or

$$C_n-C_s=\frac{T}{8\pi}\left(\frac{\partial H_0}{\partial T}\right)^2.$$

One sees that the thermodynamic properties of the ideal Bose gas are not in agreement with the properties of actual superconductors. The transition in the magnetic field is still of the second kind, having no latent heat; however, a jump in the specific heat occurs for $H\neq 0$.[24]

There now remains to calculate $M(H')$ from the basic equations (5.5) and (5.6). We shall do this separately for the absolute zero and for higher temperatures.

(a) *Absolute Zero: $\mu_0 B\gg kT$*

Here, (5.5) and (5.6) reduce to

$$\omega=-n\mu_0H'\left(\frac{T}{T_c}\right)^{\frac{3}{2}}\frac{2}{\zeta(\frac{3}{2})}\sum_{\nu=1}^{\infty}\frac{\exp[-\alpha(\mu_0H'-\mu)\nu]}{\nu^{\frac{3}{2}}}, \tag{5.18}$$

$$n=n\left(\frac{T}{T_c}\right)^{\frac{3}{2}}\frac{2}{\zeta(\frac{3}{2})}\alpha\mu_0H'\sum_{\nu=1}^{\infty}\frac{\exp[-\alpha(\mu_0H'-\mu)\nu]}{\nu^{\frac{1}{2}}}, \tag{5.19}$$

[22] C. J. Gorter and H. B. G. Casimir, Physica **1**, 305 (1934).

[23] See reference 9, paragraph 2.

[24] It should be borne in mind that (5.11) and all subsequent equations become invalid near the transition temperature. It follows from (5.17) that $H_0(T)$ must have a vanishing derivative at $T=T_c$, since there $C_n=C_s=0$.

where use has been made of the equation valid for $T<T_c$.

$$n_0=n(T/T_c)^{\frac{3}{2}}, \tag{5.20}$$

The factor $(T/T_c)^{\frac{3}{2}}\alpha\mu_0 H$ tends to zero as $T\to 0$, and therefore the sum in (5.19) must tend to infinity as $T\to 0$, i.e.,

$$\alpha(\mu_0 H'-\mu)\to\infty \quad \text{as} \quad T\to 0.$$

At $T=0$ therefore, one has

$$\mu=\mu_0 H',$$
$$\omega=0,$$

and thus the free energy per unit volume is

$$\varphi(H',0)\equiv\omega+\mu n=n\mu_0 H'. \tag{5.21}$$

The magnetization is

$$M=-\partial\varphi/\partial H'=-n\mu_0. \tag{5.22}$$

It is of no great use to make an expansion in powers of temperature starting from (5.18) and (5.19), because the range of validity of such an expansion is limited to a very small interval, *viz.*:

$$T/T_c\ll\eta'\equiv 4\pi n\mu_0^2/kT_c, \tag{5.23}$$

and this latter ratio is of the same order of magnitude as η (5.9) and thus very small for all reasonable values of e, m, n.

(b) Finite Temperatures: $\mu_0 H\ll kT$

For $\alpha\mu_0 H\ll 1$, it is shown in Appendix III that the expressions (5.5) and (5.6) can be transformed into

$$n-n_0=n_0\frac{2\sqrt{\pi}}{\zeta(\frac{3}{2})}(\alpha\mu_0 H')^{\frac{1}{2}}F(s), \tag{5.22}$$

$$\omega+\mu n-\omega_0=n_0\frac{2\sqrt{\pi}}{\zeta(\frac{3}{2})}kT(\alpha\mu_0 H')^{\frac{3}{2}}\Phi(s), \tag{5.23}$$

where n_0 is given by (5.7), ω_0 by (5.5) with $H'=0$, $\mu=0$, $s=1-\mu/(\mu_0 H')$,

$$\Phi(s)=\pi^{-\frac{1}{2}}G(s)+(1-s)F(s), \tag{5.24}$$

and

$$F(s)=\sum_{n=1}^{\infty}\frac{\cos[\pi(ns-\frac{1}{4})]}{n^{\frac{1}{2}}},$$
$$G(s)=\sum_{n=1}^{\infty}\frac{\sin[\pi(ns-\frac{1}{4})]}{n^{\frac{1}{2}}}. \tag{5.25}$$

An expansion of these functions for small $|s|$ can be obtained by using the well-known expansion[25]

$$Z_\alpha(x)\equiv\sum_{\nu=1}^{\infty}\frac{e^{-\nu x}}{\nu^\alpha}=\frac{\pi}{\sin\pi\alpha}\frac{x^{\alpha-1}}{\Gamma(\alpha)}+\sum_{\lambda=0}^{\infty}\frac{(-1)^\lambda}{\lambda!}\zeta(\alpha-\lambda)x^\lambda \tag{5.26}$$

valid for nonintegral α, $\mathrm{Re}(x)>0$, and $|x|<2\pi$. One finds

$$F(s)=s^{-\frac{1}{2}}+\frac{\zeta(\frac{1}{2})}{\sqrt{2}}+\frac{\pi\zeta(-\frac{1}{2})}{\sqrt{2}}s+O(s^2),$$
$$G(s)=-\frac{\zeta(\frac{3}{2})}{\sqrt{2}}+2\pi s^{\frac{1}{2}}+\frac{\pi\zeta(\frac{1}{2})}{\sqrt{2}}s+O(s^2). \tag{5.27}$$

Introducing the dimensionless variables

$$h=H'/4\pi n\mu_0^2, \quad t=T/T_c$$

(5.24) and (5.25) read

$$1=\frac{2\sqrt{\pi}}{\zeta(\frac{3}{2})}\frac{t}{1-t^{\frac{1}{2}}}(h\eta')^{\frac{1}{2}}F(s), \tag{5.30}$$

$$\frac{\varphi(H',T)-\varphi(0,T)}{4\pi n^2\mu_0^2}=\frac{2\sqrt{\pi}}{\zeta(\frac{3}{2})}t(\eta')^{\frac{1}{2}}h^{\frac{3}{2}}\Phi(s). \tag{5.31}$$

From (5.30) it follows that for $\eta' h\ll 1-t^{\frac{1}{2}}$ we can restrict ourselves to the first terms in the expansion (5.29), so that $\Phi(s)\approx F(s)$. Then, (5.31) yields

$$\varphi(H',T)-\varphi(0,T)=4\pi n^2\mu_0^2(1-t^{\frac{1}{2}})h=n\mu_0(1-t^{\frac{1}{2}})H', \tag{5.32}$$

and

$$M(H',T)\equiv-\frac{\partial\varphi(H',T)}{\partial H'}=-n\mu_0(1-t^{\frac{1}{2}}),$$
$$M(H',T)=-n_s\mu_0, \tag{5.33}$$

which proves our earlier statement. (5.33) is valid as long as

$$1-(T/T_c)^{\frac{1}{2}}\gg(\eta')^{\frac{1}{2}}. \tag{5.34}$$

6. CONCLUSION AND OUTLOOK

As shown in the preceding sections, an ideal gas of charged bosons at low temperatures exhibits qualitatively the same magnetic and thermodynamic properties as a superconductor. The only differences are the continuity of the specific heat through the transition point and the form of the $B(H)$ curve. The nonequilibrium properties (e.g., the vanishing of electrical resistance) are clearly outside the scope of the model, since it does not provide any mechanism for electrical resistance even above the transition temperature.

In view of the fact that the Bose gas is the only known system with superconducting properties, it is tempting to assume that the occurrence of superconductivity in metals is due to the formation of some kind of charged bosons at low temperatures. Without prejudicing the nature of these bosons, one may get some insight about them by trying to fit experimental data to the ideal-gas model.

The model contains three parameters, the mass m, the charge e and the density n of the bosons. These can be fitted to three experimental data: the transition temperature T_c, the penetration depth d_0 at $T=0$, and the

[25] J. E. Robinson, Phys. Rev. **83**, 678 (1951); M. R. Schafroth, Proc. Phys. Soc. (London) **A67**, 33 (1953); J. Clunie, Proc. Phys. Soc. (London) **A67**, 682 (1954).

critical field $H_c(0)$ at $T=0$. If one tries this, one gets utterly implausible results ($n\approx 10^{17}$ cm^{-3}, $m\approx 10^{8}$ electron masses). Only the charge falls into the reasonable order of magnitude of a few elementary charges. This fact may give a clue for the understanding of the actual situation if one notices that the charge is determined through (5.14) by $H_c(0)$ and d_0 alone, i.e., by two quantities at zero temperature, whereas m and n require comparison of properties at $T=0$ and $T=T_c$. It seems, therefore, that one will have to consider a creation of more bosons with decreasing temperature. If one allows for this and assumes a reasonable boson mass of the order of the electron mass, one can compute the density n of bosons at $T=T_c$ from (1.1) and at $T=0$ from (3.7), and one finds

$$n(T)_c\approx 10^{17}\ \text{cm}^{-3},$$
$$n(0)\approx 10^{22}\ \text{cm}^{-3},$$

i.e., a considerable increase. On the other hand, this picture does not affect the relation (5.14) between $H_c(0)$ and d_0 which shows that the charge of the bosons is of the order of a few electron charges e_0. For tin, e.g., taking $d_0=5.2\times 10^{-6}$ cm and $H_c(0)=300$ gauss one gets $e=4e_0$.

This picture has the further consequence that η' (5.23) and η (5.9) are no longer of the same order of magnitude, since the former now involves $n(0)/[n(T_c)]^{\frac{1}{3}}$, and the latter $[n(T_c)]^{\frac{1}{3}}$. The temperature interval in which $K_n(q^2)$ is important and (5.11) fails, thus remains small ($(T_c-T)/T_c\gg 10^{-7}$), whereas $\eta'=\mu_0 H_c(0)/kT_c\sim 10^{-2}$.

It is also interesting to note that the relation (5.14), as a function of temperature, holds approximately in actual superconductors, the product of d^2H_0 varying only by about 50% over the range $0<T<T_c$, whereas the temperature dependence (3.6) of d alone is quite different from the one in superconducting metals. This seems to corroborate the assumption that actual superconductivity is related to the properties of a Bose gas whose total particle number depends on temperature.

A theoretical picture which accounts for a temperature-dependent density of bosons has been proposed by the author[11] on the assumption that the bosons are resonant two-electron states. However, a rough treatment which neglected the width of the resonance altogether was not sufficient to fit the facts quantitatively, mainly because the increase in the density of bosons between the transition point and absolute zero could not be made large enough. It is, however, expected that a more refined elaboration of the same picture would lead to appreciable improvement.

The author is deeply indebted to Dr. J. M. Blatt and Dr. S. T. Butler for numerous valuable discussions and criticisms. Their stimulating influence was vital to the development of this work. His thanks are also due to Professor W. Pauli for a very interesting correspondence over this work which reshaped important parts of it. Helpful criticism by Professor H. Froehlich is also gratefully acknowledged.

APPENDIX I. EXPANSION OF $k(z)$

We wish to expand

$$k_\epsilon(z)=\int_0^\infty dx\left[\frac{x}{2z}(x^2-z^2)\times\log\left|\frac{x-z}{x+z}\right|+x^2\right]\frac{1}{\exp(x^2+\epsilon)-1} \tag{I.1}$$

in powers of z in the limit $\epsilon\to 0$. First we notice that the integrand is even in x, so that

$$k_\epsilon(z)=\tfrac{1}{2}\int_{-\infty}^{+\infty} dx\left[\frac{x}{2z}(x^2-z^2)\times\log\left|\frac{x-z}{x+z}\right|+x^2\right]\frac{1}{\exp(x^2+\epsilon)-1}. \tag{I.1'}$$

This can be expressed in the complex x-plane, with a cut from $-z$ to $+z$, in which $\log(x-z)$ is real for real $x-z>0$.

$$k_\epsilon(z)=\tfrac{1}{2}\,\mathrm{Re}\int_C dx\left[\frac{x}{2z}(x^2-z^2)\times\log\left(\frac{x-z}{x+z}\right)+x^2\right]\frac{1}{\exp(x^2+\epsilon)-1} \tag{I.2}$$

where Re denotes "real part" and C is the path shown in Fig. 1. (2) can then be reduced to an integral over a path C' on which everywhere $|x|>z$ and a residue at the pole $x_0=+i\epsilon^{\frac{1}{2}}$.

The residue contributes

$$R(\epsilon,z)=\tfrac{1}{2}\,\mathrm{Re}\left\{2\pi i\ \underset{x=x_0}{\text{residue}}\left(\left[\frac{x}{2z}(x^2-z^2)\times\log\left(\frac{x-z}{x+z}\right)+x^2\right)\frac{1}{\exp(x^2+\epsilon)-1}\right\}$$
$$=+\pi\,\mathrm{Re}\left\{i\left[\frac{i\epsilon^{\frac{1}{2}}}{2z}(-\epsilon-z^2)\times\log\left(\frac{i\epsilon^{\frac{1}{2}}-z}{i\epsilon^{\frac{1}{2}}+z}\right)-\epsilon\right]\frac{1}{2i\epsilon^{\frac{1}{2}}}\right\}. \tag{I.3}$$

In the limit $\epsilon\to 0$, we have

$$\log\left(\frac{i\epsilon^{\frac{1}{2}}-z}{i\epsilon^{\frac{1}{2}}+z}\right)\to i\pi,$$

and thus

$$R(0,z)=(\pi^2/4)z. \tag{I.4}$$

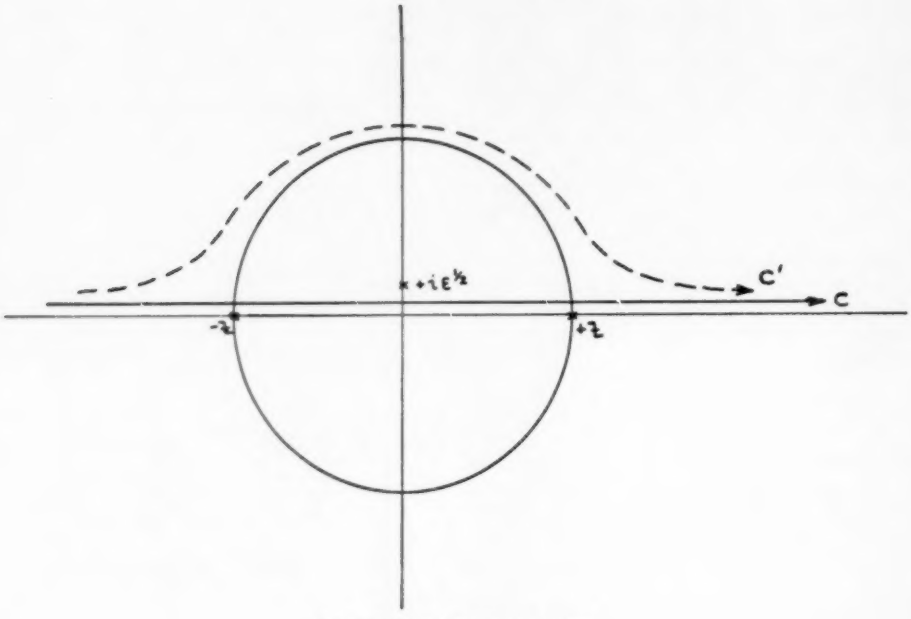


FIG. 1. Paths in the complex x-plane.

We therefore get

$$k_0(z)=\frac{\pi^2}{4}z+\tfrac{1}{2}\,\mathrm{Re}\int_{C'}dx\left[\frac{x}{2z}(x^2-z^2)\times\log\left(\frac{x-z}{x+z}\right)+x^2\right]\frac{1}{\exp(x^2)-1}. \quad \text{(I.5)}$$

In the remaining integral, $|z|<|x|$, so that we can expand the expression in brackets in powers of z/x. The first term is easily calculated to be

$$\left[\frac{x}{2z}(x^2-z^2)\log\left(\frac{x-z}{x+z}\right)+x^2\right]=\tfrac{2}{3}z^2+O(z^4), \quad \text{(I.6)}$$

so that

$$k_0(z)=\frac{\pi^2}{4}z+\frac{z^2}{3}\,\mathrm{Re}\int_{C'}dx\frac{1}{\exp(x^2)-1}. \quad \text{(I.7)}$$

The remaining z-independent integral can be reduced by the substitution $x^2=t$, to the well-known integral representation of Riemann's zeta function[26] $\zeta(s)$ for $s=\frac{1}{2}$ and we finally get

$$k_0(z)=\frac{\pi^2}{4}z+\frac{\sqrt{\pi}}{3}\zeta(\tfrac{1}{2})z^2+O(z^4), \quad \text{(I.8)}$$

which is the result (2.29).

[26] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, Cambridge, 1946), fourth edition, p. 266.

APPENDIX II. ESTIMATE OF RADII OF ORBITS IN MAGNETIC FIELD

The radius of the classical orbit of a particle moving with energy E in a field H' is given by

$$\rho=(c/eH')(2mE)^{\frac{1}{2}}. \quad \text{(II.1)}$$

As long as $\mu_0H'\ll kT$, we may replace E by the average kinetic energy of a free particle, i.e.,

$$(2mE)^{\frac{1}{2}}\approx\frac{1}{n}\frac{m^2}{\pi^2\hbar^3}(kT)^2\zeta(2). \quad \text{(II.2)}$$

The maximum value for H' we need is

$$H_0(0)=4\pi n\mu_0. \quad \text{(II.3)}$$

Therefore

$$\frac{\rho}{a}\geq\mathrm{const}\frac{a}{r_0}\left(\frac{T}{T_c}\right)^2=\frac{\mathrm{const}}{\eta}\left(\frac{T}{T_c}\right)^2, \quad \text{(II.4)}$$

where $r_0=e^2/mc^2$, $a=n^{-\frac{1}{3}}=$ mean distance between particles.

This estimate breaks down near $T=0$. Here we have to assume $\mu_0H'\gg kT$, and to take for E the ground-state energy of a particle in a magnetic field:

$$E\approx\mu_0H', \quad \text{(II.5)}$$

(II.1) then yields

$$\rho\geq d_0\gg a, \quad \text{(II.6)}$$

where d_0 is the penetration depth at absolute zero, (5.7).

(II.4) and (II.6) establish the required estimate

$$\rho/a \gg 1 \tag{II.7}$$

for the whole range of temperature and fields of interest.

APPENDIX III. REDUCTION OF (5.5) AND (5.6) TO (5.24) AND (5.25)

In the case of $\alpha\mu_0 H' \ll 1$, it is convenient to consider instead of ω and n the following linear combinations

$$\zeta(\tfrac{3}{2})\frac{n_0-n}{n_0} \equiv A = \sum_{\nu=1}^{\infty} \frac{1}{\nu^{\frac{1}{2}}}\left[1-e^{\alpha\mu\nu}\frac{x\nu}{\sinh(x\nu)}\right], \tag{III.1}$$

$$\alpha\zeta(\tfrac{3}{2})\frac{\omega+\mu n-\omega_0}{n_0} \equiv B = \sum_{\nu=1}^{\infty} \frac{1}{\nu^{5/2}} \times\left[1-e^{\alpha\mu\nu}\frac{x\nu}{\sinh(x\nu)}(1-\alpha\mu\nu)\right], \tag{III.2}$$

with $x \equiv \alpha\mu_0 H' \ll 1$. These are the physical quantities actually required and are also mathematically convenient for the following reason: Due to our assumption $x \ll 1$ and the ensuing one $\alpha\mu \ll 1$ (in view of $0 < \mu < \mu_0 H'$), we can replace the sums over ν by integrals, provided these converge at the lower limit. The combinations (III.1) and (III.2) are just such that this holds, and we therefore write

$$A = \int_0^{\infty} \frac{d\nu}{\nu^{\frac{1}{2}}}\left[1-e^{\alpha\mu\nu}\frac{x\nu}{\sinh(x\nu)}\right], \tag{III.3}$$

$$B = \int_0^{\infty} \frac{d\nu}{\nu^{5/2}}\left[1-e^{\alpha\mu\nu}\frac{x\nu}{\sinh(x\nu)}(1-\alpha\mu\nu)\right]. \tag{III.4}$$

These integrals can be evaluated by expanding their integrands into partial fractions, using the theorem of Mittag-Leffler. One finds that

$$\frac{1}{\nu}\left[1-e^{\alpha\mu\nu}\frac{x\nu}{\sinh(x\nu)}\right] = \sum_{n=1}^{\infty}(-1)^{n-1}\left\{\frac{\exp\left(i\pi n\frac{\alpha\mu}{x}\right)}{\nu-\frac{i\pi n}{x}} + \frac{\exp\left(-i\pi n\frac{\alpha\mu}{x}\right)}{\nu+\frac{i\pi n}{x}}\right\}, \tag{III.5}$$

$$\frac{1}{\nu^2}\left[1-e^{\alpha\mu\nu}\frac{x\nu}{\sinh(x\nu)}(1-\alpha\mu\nu)\right] = \sum_{n=1}^{\infty}(-1)^{n-1}\left\{\frac{x}{n\pi i}\left(1-in\pi\frac{\alpha\mu}{x}\right)\frac{\exp\left(i\pi n\frac{\alpha\mu}{x}\right)}{\nu-\frac{i\pi n}{x}} - \frac{x}{n\pi i}\left(1+in\pi\frac{\alpha\mu}{x}\right)\frac{\exp\left(-i\pi n\frac{\alpha\mu}{x}\right)}{\nu+\frac{i\pi n}{x}}\right\}. \tag{III.6}$$

The integration in A and B then gives

$$A = 2(\pi x)^{\frac{1}{2}}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^{\frac{1}{2}}}\cos\left[\pi\left(n\frac{\alpha\mu}{x}+\tfrac{1}{4}\right)\right], \tag{III.7}$$

$$B = \frac{2}{\sqrt{\pi}}x^{\frac{1}{2}}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^{\frac{1}{2}}}\sin\left[\pi\left(n\frac{\alpha\mu}{x}+\tfrac{1}{4}\right)\right] - 2(\pi x)^{\frac{1}{2}}\alpha\mu\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^{\frac{1}{2}}}\cos\left[\pi\left(n\frac{\alpha\mu}{x}+\tfrac{1}{4}\right)\right]. \tag{III.8}$$

Putting $s = 1-\alpha\mu/x = 1-\mu/(\mu_0 H')$ these expressions become

$$A = -2\sqrt{\pi}(\alpha\mu_0 H')^{\frac{1}{2}}\sum_{n=1}^{\infty}\frac{1}{n^{\frac{1}{2}}}\cos[\pi(ns-\tfrac{1}{4})] = -2\sqrt{\pi}(\alpha\mu_0 H')^{\frac{1}{2}}F(s). \tag{III.9}$$

$$B = \frac{2}{\sqrt{\pi}}(\alpha\mu_0 H')^{\frac{1}{2}}\sum_{n=1}^{\infty}\frac{1}{n^{\frac{1}{2}}}\sin[\pi(ns-\tfrac{1}{4})] + 2(\pi)^{\frac{1}{2}}(\alpha\mu_0 H')^{\frac{1}{2}}\alpha\mu\sum_{n=1}^{\infty}\frac{1}{n^{\frac{1}{2}}}\cos[\pi(ns-\tfrac{1}{4})] = \frac{2}{\sqrt{\pi}}(\alpha\mu_0 H')^{\frac{1}{2}}\{G(s)+\pi(1-s)F(s)\}, \tag{III.10}$$

with $F(s)$ and $G(s)$ defined by (5.27). Inserting (III.9) and (III.10) into (III.1) and (III.2) readily gives (5.24) and (5.25).

Superfluidity of an Ideal Bose-Einstein Gas

J. M. Blatt and S. T. Butler
The F. B. S. Falkiner Nuclear Research and Adolph Basser Computing Laboratories, School of Physics, The University of Sydney, Sydney, Australia*

(Received April 4, 1955)

An equilibrium superfluid is defined as a substance whose moment of inertia in thermal equilibrium is less than the classical value. The ideal Bose-Einstein gas below its transition temperature is shown to be an equilibrium superfluid. However, a more detailed analysis of the relation between angular momentum and angular velocity shows that experiments carried out with practical angular momenta and velocities would give an abnormally *large* apparent moment of inertia in thermal equilibrium, i.e., the gas would appear to be an "infrafluid." Nevertheless, a weakly interacting Bose-Einstein gas can be used as a model of a superfluid, if the superfluid state is assumed to be metastable rather than the thermodynamic equilibrium state.

1. INTRODUCTION

LONDON[1] first proposed that the peculiar phase transition (λ point) in liquid helium at 2.186°K is essentially the same as the condensation phenomenon exhibited by an ideal Bose-Einstein gas,[2,3] only slightly disturbed by the interactions between the particles. It is the purpose of this paper to study the properties of an ideal Bose-Einstein gas below the λ point when the vessel in which the gas is contained is subjected to a uniform rotation. This "rotating bucket" experiment was first discussed by Landau[4] on the basis of his theory of the superfluidity of liquid helium. Landau predicted that liquid helium would show an abnormal relation between angular momentum L and angular velocity ω in *statistical equilibrium*. The moment of inertia $I=\partial L/\partial\omega$ would be less than its classical value I_0. Subsequent experimental work by Andronikashvili[5] seemed to bear out this prediction.

We shall define the equilibrium moment of inertia of a substance as

$$I=\lim_{\omega\to 0}\left[\frac{\partial \bar{L}(\omega)}{\partial\omega}\right], \tag{1.1}$$

where $\bar{L}(\omega)$ is the *equilibrium* value of the angular momentum of the substance in a container rotating uniformly with angular velocity ω. We shall further define an *equilibrium superfluid* as a substance for which I is less than the classical moment of inertia I_0, the latter being computed in the usual way; in a cylindrical vessel of radius R, filled with N particles each of mass M, I_0 is given by

$$I_0=\tfrac{1}{2}NMR^2. \tag{1.2}$$

There are 3 different ways of defining equilibrium superfluidity, in increasing order of stringency of the requirement:

(1) The initial slope of the curve of $\bar{L}$ *vs* ω, evaluated at $\omega=0$, is less than the classical value I_0. This is our requirement.

(2) $\bar{L}$ maintains this nonclassical slope for a finite range of values of the circumferential velocity $v=R\omega$; however, this range may become smaller and smaller as the radius R of the vessel increases. This is the suggestion of London[6] who expects an upper limit to superfluid flow of order $v=R\omega\lesssim\hbar/MR$.

(3) $\bar{L}$ maintains the nonclassical slope for values of the circumferential velocity less than some "critical velocity," this latter being independent of the radius of the vessel. This (most stringent) requirement is assumed to be satisfied in the usual two-fluid model of liquid helium.

This paper proves that the ideal Bose-Einstein gas below its condensation point is an equilibrium superfluid in the sense of the first and second requirements above, but does not satisfy the third requirement. The nonclassical slope of $\bar{L}$ is maintained until the angular velocity ω reaches the value ω_1 given by

$$\omega_1=4.45\hbar/MR^2. \tag{1.3}$$

Thereafter, however, the picture changes suddenly. At $\omega=\omega_1$, the angular momentum $\bar{L}(\omega)$ has a discontinuity, due to the fact that the condensed particles shift suddenly from the state with angular momentum quantum number $m=0$ to the state with $m=1$. Thermodynamically, this is a phase transition of the first kind, with a latent heat. Similar phase transitions occur at higher values of ω. We shall show that a conventional measurement of the moment of inertia of an ideal Bose-Einstein gas in thermal equilibrium would yield an averaged value I_{eff} which is larger than the classical value I_0, i.e., the gas would appear to be an "infrafluid" rather than a superfluid. Thus, although the ideal Bose-Einstein gas is an equilibrium superfluid in the sense of London,[1,6] it fails to exhibit the kind of equilibrium superfluidity postulated by Landau[4] and Tisza.[7]

* Also supported by the Nuclear Research Foundation within the University of Sydney.

[1] F. London, Phys. Rev. **54**, 947 (1938).
[2] A. Einstein, Ber. Berl. Akad. **261** (1924); 3 (1925).
[3] B. Kahn and G. E. Uhlebeck, Physica **5**, 399 (1938).
[4] L. Landau, J. Phys. (U.S.S.R.) **5**, 71 (1941).
[5] E. L. Andronikashvili, J. Phys. (U.S.S.R.) **10**, 201 (1946).
[6] H. London, Phys. Soc. Camb. Conference Report, **48**, 1946 (Physical Society, London, 1947).
[7] L. Tisza, Phys. Rev. **72**, 838 (1947).

Nevertheless, the Bose-Einstein gas can be used as a model of a superfluid if we drop the requirement of thermal equilibrium; the nonequilibrium superfluid behavior which results is highly suggestive, and we believe contains the essence of the superfluid properties observed in actual liquid helium.

2. MOMENT OF INERTIA OF AN IDEAL BOSE-EINSTEIN GAS

The Hamiltonian of an ideal Bose-Einstein gas in a stationary container is

$$H_0=\sum_{i=1}^{N} p_i^2/2M=\sum_{i=1}^{N} h_{0i}. \tag{2.1}$$

In order to discuss thermal equilibrium of the gas inside a rotating container, it is necessary[8] to use a system of coordinates rotating along with the container. The operator for the z component of the angular momentum of particle number i is

$$l_i=-i\hbar\left(x_i\frac{\partial}{\partial y_i}-y_i\frac{\partial}{\partial x_i}\right), \tag{2.2}$$

where x_i, y_i, z_i are the coordinates of particle i measured in the rotating frame of reference. It can be shown, however,[8] that the expectation value of l_i is the angular momentum of particle i with respect to the *stationary* frame of reference. The z-component of the total angular momentum of all the particles will be denoted by L:

$$L=\sum_{i=1}^{N} l_i. \tag{2.3}$$

The canonical transformation to the rotating system of coordinates leads to the new Hamiltonian

$$H'=\sum_{i=1}^{N}(h_{0i}-\omega l_i)=H_0-\omega L. \tag{2.4}$$

This Hamiltonian determines the properties of the gas in statistical equilibrium within the rotating container, in the usual way. We shall be particularly interested in the statistical average value of L, given by

$$\bar{L}(\omega)=\frac{\mathrm{Trace}[L\exp(-\beta H')]}{\mathrm{Trace}[\exp(-\beta H')]}, \tag{2.5}$$

where $\beta=(kT)^{-1}$, and the trace is to be taken over all permissible quantum states of the N-particle system (all states with symmetric wave functions for Bose-Einstein statistics).

Since there are no interactions between the particles, the calculation of $\bar{L}$ can be simplified. We define the symbol "trace" to mean a trace over all quantum states of a single particle in the container (rather than over quantum states of the full N-particle system). The chemical potential μ is then defined by the relation:

$$\mathrm{trace}\left\{\frac{1}{\exp[\beta(h_0-\omega l-\mu)]-1}\right\}=N, \tag{2.6}$$

where N is the total number of particles. By reflection symmetry ($\theta'=-\theta$) μ is a symmetric function of ω. The equilibrium value of L, (2.5) is then given by

$$\bar{L}(\omega)=\mathrm{trace}\left\{l\frac{1}{\exp[\beta(h_0-\omega l-\mu]-1}\right\}. \tag{2.7}$$

In a noncylindrical container, h_0 and l fail to commute, and so do H_0 and L for the N-particle system. In the N-particle formulation (2.5) the error made by assuming that H_0 and L commute is of order N^{-1} and hence unimportant. The corresponding error in the one-particle formulation (2.7) is very significant for temperatures below the condensation point of the Bose-Einstein gas. Since these complications add nothing to the argument, we shall consider cylindrically symmetric containers.

We label the common eigenstates of h_0 and l by the two indices m and γ, where $\hbar m$ is the eigenvalue of l and γ stands for the two additional quantum numbers necessary to specify the quantum state uniquely. The eigenvalue of h_0 in the quantum state (m,γ) will be called $\epsilon_{m\gamma}$; the lowest $\epsilon_{m\gamma}$ for a given value of m will be called ϵ_{m0}.

We first evaluate the initial slope of $\bar{L}$ as a function of ω; that is, we differentiate (2.7) with respect to ω and then set $\omega=0$. In the range of ω for which the differentiation is permitted, the chemical potential μ is also a differentiable function of ω; furthermore, since μ is an even function of ω, the derivative $\partial\mu/\partial\omega$ vanishes at $\omega=0$. We therefore obtain the following expression for the moment of inertia I:

$$I=\left(\frac{d\bar{L}}{d\omega}\right)_{\omega=0}=\beta\,\mathrm{trace}\left\{l^2\frac{\exp[\beta(h_0-\mu_0)]}{(\exp[\beta(h_0-\mu_0)]-1)^2}\right\}$$

$$=\hbar^2\beta\sum_{\gamma}\sum_{m=-\infty}^{\infty} m^2\frac{\exp[\beta(\epsilon_{m\gamma}-\mu_0)]}{\{\exp[\beta(\epsilon_{m\gamma}-\mu_0)]-1\}^2}. \tag{2.8}$$

The sum in (2.8) is rather awkward to handle since $\epsilon_{m\gamma}$ is a complicated expression for a cylindrical container. However, we can simplify the evaluation of (2.8) tremendously provided that the main contribution to the trace comes from quantum states with de Broglie wavelengths much smaller than the dimensions of the bucket. This is obviously so above the Bose-Einstein condensation point; it is also true below the condensation point, because the ground state of the single particle spectrum is the state ϵ_{00}, which has $m=0$ and therefore contributes nothing to (2.8). If the de Broglie

[8] For detailed arguments see the following paper, Blatt, Butler, and Schafroth, Phys. Rev. **100**, 481 (1955).

wavelength is small compared to the linear dimensions of the bucket, the boundary condition $\psi=0$ at the wall of the container is unimportant. We can work with plane wave functions $\exp(i\mathbf{p}\cdot\mathbf{r}/\hbar)$ which form a complete set, but fail to satisfy the boundary condition. The error is of order $\lambda/\mathfrak{L}$ where λ is the de Broglie wavelength and $\mathfrak{L}$ is the linear dimension of the container. We then get

$$I=\beta\sum_{p}{}'\langle\mathbf{p}|l^2|\mathbf{p}\rangle\frac{\exp[\beta(p^2/2M-\mu_0)]}{\{\exp[\beta(p^2/2M-\mu_0)]-1\}^2}, \quad (2.9)$$

where the prime on the sum indicates that very small values of p ($\hbar/p\sim\mathfrak{L}$) are to be excluded. The diagonal matrix element of the operator l^2 in a plane wave state is given by (V = volume of the container)

$$\langle\mathbf{p}|l^2|\mathbf{p}\rangle=V^{-1}\int(xp_y-yp_x)^2d^3\tau$$
$$=\langle x^2\rangle p_y^2+\langle y^2\rangle p_x^2-2\langle xy\rangle p_xp_y. \quad (2.10)$$

The volume integral in (2.10) is over the interior of the container. For a cylindrical container of radius R we get

$$\langle x^2\rangle=\langle y^2\rangle=R^2/4 \quad \langle xy\rangle=0. \quad (2.11)$$

However, this part of our argument applies to a container of arbitrary shape. The shape of the container becomes important only when we consider quantum states with de Broglie wavelengths comparable to the linear dimensions of the container.

We substitute (2.10) into (2.9) and replace the sum over quantum states $\mathbf{p}$ by an integral in the conventional way. Consider the term proportional to $\langle x^2\rangle$; this term can be transformed by an integration by parts:

$$-\beta V(2\pi)^{-3}\langle x^2\rangle\int d^3p\,p_y(M/\beta)\frac{\partial}{\partial p_y}$$
$$\times\{\exp[\beta(p^2/2M-\mu_0)]-1\}^{-1}=M\langle x^2\rangle V(2\pi)^{-3}\int d^3p$$
$$\times\{\exp[\beta(p^2/2M-\mu_0)]-1\}^{-1}. \quad (2.12)$$

As a result of the factor p^2 in the volume element d^3p, states with very small values of p make no significant contribution to the last integral, even if μ_0 is very close to zero. Thus the integral correctly represents the primed sum (2.9).

Above the condensation point, the second integral in (2.12) represents the trace (2.6), i.e., we get the expected result for the moment of inertia:

$$I=I_0=NM\langle x^2+y^2\rangle \quad T>T_\lambda. \quad (2.13)$$

Below the condensation point, however, we get a different result: the integral in (2.12) still represents the primed sum (2.9), and hence the trace (2.8); but it is no longer simply related to the trace (2.6), because the ground state ϵ_{00} contributes appreciably to(2.6) but is excluded from (2.12). The integral on the right side of (2.12) represents the number of noncondensed particles, N'. The well-known expression for N' is

$$N'=N(T/T_\lambda)^{\frac{3}{2}}, \quad (2.14)$$

and hence the moment of inertia of the ideal Bose-Einstein gas below the condensation point is

$$I=N'M\langle x^2+y^2\rangle=(T/T_\lambda)^{\frac{3}{2}}I_0. \quad (2.15)$$

This concludes the proof that the ideal Bose-Einstein gas *in thermal equilibrium* has a moment of inertia less than the classically expected value, i.e., the ideal Bose-Einstein gas is an equilibrium superfluid. As far as we know, this is the only system known which exhibits equilibrium superfluidity.

There have been many attempts in the literature to derive equilibrium superfluidity from the assumption of an energy gap between the ground state of the N-particle system and some or all of the excited states. For example, Landau[4] assumes that the only states contiguous to the ground state in energy are phonon states in which the circulation vector $\nabla\times\mathbf{v}$ vanishes. However, no one has been able to demonstrate the existence of such an energy gap from first principles. Thus the only system known at present to be an equilibrium superfluid does *not* have an energy gap.

3. ANGULAR MOMENTUM OF AN IDEAL BOSE-EINSTEIN GAS

The definition (1.1) of the moment of inertia I in thermal equilibrium is reasonable from a theoretical point of view. Experimentally, however, the angular velocity ω cannot be made smaller than some minimum value, determined by the sensitivity of the measuring equipment. Thus the experimentalist determines not the derivative $\partial\bar{L}/\partial\omega$ but rather a difference quotient

$$I_{\text{eff}}=\Delta\bar{L}/\Delta\omega. \quad (3.1)$$

We now propose to study the angular momentum $\bar{L}(\omega)$ of the ideal Bose-Einstein gas in thermal equilibrium as a function of the angular velocity ω, rather than merely the derivative $\partial\bar{L}/\partial\omega$ at $\omega=0$. This will enable us to find the quantity I_{eff}, of (3.1).

According to (2.7), $\bar{L}(\omega)$ is given by

$$\bar{L}(\omega)=\sum_{\gamma}\sum_{m=-\infty}^{\infty}\frac{\hbar m}{\exp[\beta(\epsilon_{m\gamma}-\hbar\omega m-\mu)]-1}. \quad (3.2)$$

Let us now see what happens as ω is increased slowly from zero. Initially, the macroscopic number of particles in state ϵ_{00} does not contribute to (3.2), since $m=0$ for these particles. Thus initially $\bar{L}$ is a linear function of ω with the nonclassical slope (2.15).

However, this behavior does not persist. As ω increases smoothly, a point is reached at which the lowest energy level with $m=1$, i.e., ϵ_{10}, fulfills the equation

$$\epsilon_{10}-\hbar\omega=\epsilon_{00}. \quad (3.3)$$

At this value of ω the thermal equilibrium distribution changes abruptly. A macroscopic number of particles

shifts from state ϵ_{00} to state ϵ_{10}. Correspondingly, there occurs a finite jump in the value of the equilibrium angular momentum $\bar{L}$; the discontinuity in $\bar{L}$ is equal to $N_0\hbar$ where $N_0=N-N'$ is the number of condensed particles. There is also a discontinuous shift in the internal energy of the system.

Thus, thermodynamically speaking, we have a *phase transition of the first kind*.[9] The chemical potential μ is continuous with discontinuous derivative $\partial\mu/\partial\omega$; the same is true of the free energy. The internal energy and the angular momentum have discontinuities, and there is therefore a latent heat.

Let us determine the value of ω at which the transition occurs. The energy ϵ_{m0} of the lowest state with angular momentum quantum number m is given by

$$\epsilon_{m0}=\hbar^2x_m^2/2MR^2 \tag{3.4}$$

where x_m is the first zero of the Bessel function $J_m(x)$. We therefore get

$$\omega_1=(\epsilon_{10}-\epsilon_{00})/\hbar=4.45\hbar/MR^2. \tag{3.5}$$

An upper limit to the angular velocities at which equilibrium superfluidity can occur, of precisely this order of magnitude, was predicted on qualitative grounds by London.[6] Thus the ideal Bose-Einstein gas is a model of an equilibrium superfluid in London's sense, as well as in the sense of our definition (which refers only to the initial slope of $\bar{L}$ as a function of ω). However, not too much stress should be laid on this point, since the ideal Bose-Einstein gas is a very artificial system, and we shall see (in subsequent papers) that real physical systems do not show equilibrium superfluidity, no matter how small the angular velocity.

Consider a vessel of radius 1 cm and height 1 cm filled with an ideal gas of particles, each of mass equal to the mass of a helium atom. Then ω_1 is of the order of 10^{-3} radian/sec, and the discontinuity in $\bar{L}$ is of the order of 10^{-4} erg-sec. It is clear that these values are much too small for most experimental work.

The phase transition at $\omega=\omega_1$ is not the only one, but is the first of an infinite sequence. The next phase transition occurs when the lowest level with $m=2$ becomes the center of condensation, i.e., when

$$\epsilon_{20}-2\hbar\omega=\epsilon_{10}-\hbar\omega. \tag{3.6}$$

In order to estimate the value of ω at which the mth phase transition occurs, we need an estimate for the first zero x_m of $J_m(x)$. We use the asymptotic formula[10]

$$x_m=m+1.856m^{\frac{1}{3}}+O(m^{-\frac{1}{3}}), \tag{3.7}$$

and the definition

$$\hbar\omega_m=\epsilon_{m0}-\epsilon_{m-1,0} \tag{3.8}$$

to get the following value for the angular velocity at which the mth phase transition occurs:

$$\omega_m=(m+2.474m^{\frac{1}{3}}-\tfrac{1}{2}+\text{order } m^{-\frac{1}{3}})(\hbar/MR^2). \tag{3.9}$$

At each critical point ω_m the angular momentum $\bar{L}$ increases suddenly by $N_0\hbar$, where N_0 is the number of condensed particles.[11] The over-all curve of $\bar{L}$ against ω is shown in Fig. 1.

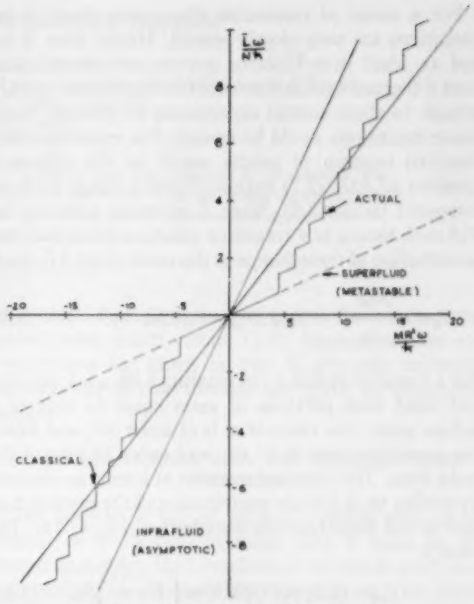


FIG. 1. Thermal equilibrium value of the angular momentum of an ideal Bose gas in a rotating container, *vs* the angular velocity of the container. The portion of the curve near $\omega=0$ has a slope less than the classically expected value, corresponding to superfluid behavior. Asymptotically, for large ω, the mean slope becomes larger than the classically expected value, corresponding to "infra-fluid" behavior, i.e., an abnormally high moment of inertia. The sharp breaks in the curve are thermodynamic transitions of the first kind, due to a macroscopic number of particles shifting from one quantum state to another. If this change-over takes a time much longer than the time in which the experiment is carried out, the angular momentum is given by the dashed curve, and exhibits superfluid behavior throughout.

[9] An objection may be raised against this terminology on the grounds that the critical value of ω, (3.3) and (3.4), depends on the radius R of the vessel, and approaches 0 as the radius of the vessel is increased. Phase transitions have a meaning only in the limit as the number of particles in the system approaches infinity, and this means a vessel of infinite volume, other things being equal. However, the volume of the vessel can be increased indefinitely without changing the radius R, merely by increasing the height. More physically, we can estimate the range $\Delta\omega$ of angular velocities over which the transition occurs for a vessel of finite height. For a typical vessel (radius = height = 1 cm) $\Delta\omega$ is of order 10^{-9} radian/sec compared to a critical value ω_1 of order 10^{-4} radian/sec. Thus the transition is already almost perfectly sharp.

[10] G. N. Watson, *Bessel Functions* (Cambridge University Press, London, 1948), p. 516.

[11] Strictly speaking, N_0 is itself a function of ω, since the number of particles which can be accommodated in the normal fluid part of the Bose-Einstein distribution changes as a result of the centrifugal force. This effect, however, can be shown to be quite negligible mathematically; the physical reason is that for typical containers (radius = 1 cm) filled with particles of the mass of a helium atom, and rotating with angular velocities of the order of 1 rad/sec, the kinetic energy due to the rotating motion, of order $MR^2\omega^2$, is many orders of magnitude less than kT (T being of order 1°K).

For a vessel of reasonable dimensions these phase transitions are very closely spaced. Hence, even if we had an ideal Bose-Einstein gas at our disposal, and even if thermal equilibrium could be established quickly enough to allow normal experiments to proceed, these phase transitions would be missed. The experimentally observed moment of inertia would be the difference quotient $\Delta\bar{L}/\Delta\omega$, (3.1) evaluated over a range $\Delta\omega$ large compared to ω_1, (3.5). Since $\bar{L}$ increases suddenly by $N_0\hbar$ each time a new transition point ω_m is reached, the contribution of these jumps to the mean slope $\Delta\bar{L}/\Delta\omega$ is

$$I_s=\frac{N_0\hbar}{\omega_m-\omega_{m-1}}=N_0MR^2(1-0.825m^{-\frac{1}{3}}+\cdots). \quad (3.10)$$

For a vessel of radius 1 cm rotating with $\omega=1$ rad/sec, and filled with particles of mass equal to that of a helium atom, the value of m is of order 10^4, and hence the correction term in (3.10) is of order 2×10^{-3} of the main term. The observed moment of inertia is obtained by adding to (3.10) the contribution of the normal fluid part of the Bose-Einstein distribution, i.e., (2.15). The total is

$$I_{\text{eff}}=(2N_0+N')\tfrac{2}{5}MR^2=\left(1+\frac{N_0}{N}\right)I_0. \quad (3.11)$$

Thus, the ideal Bose-Einstein gas is not a normal fluid even if one averages over the phase transitions; rather, it now appears to be an "infrafluid," with a moment of inertia *larger* than the classical one, in the ratio $1+(N_0/N)$.[12]

The origin of this infrafluid behavior is easy to understand: the wave function of the condensed particles is proportional to $J_m(x_{ms}r/R)$, and since x_m is the *first* zero of $J_m(z)$, this function is concentrated near the outside of the container. Unlike the normal fluid particles whose energy kT is much larger than the kinetic energy of the rotating motion, the condensed particles are pushed out immediately by the centrifugal force. Thus they contribute a moment of inertia MR^2 per particle, whereas the normal fluid component, with its uniform distribution in space, contributes $\frac{2}{5}MR^2$ per particle to the total moment of inertia. Of course, at extremely high angular velocities even the normal fluid component is pushed out toward the rim of the bucket, and the limiting value of the over-all moment of inertia per particle is MR^2. However, the ideal Bose-Einstein gas below the condensation point contains a non-negligible number of particles which are pushed out towards the rim practically immediately by the centrifugal force, long before the normal centrifugal effects are felt. Thus "infrafluidity" is not merely a particularly strong centrifugal effect understandable on a classical basis, but is a distinct quantum-mechanical effect of the Bose-Einstein statistics.

[12] The contrary statement in our earlier Letter to the Editor [J. M. Blatt and S. T. Butler, Phys. Rev. **96**, 1141 (1954)] was due to our having overlooked a factor 2.

4. IDEAL BOSE-EINSTEIN GAS AS A MODEL OF A NONEQUILIBRIUM SUPERFLUID

The preceding section has shown that the ideal Bose-Einstein gas, in spite of its nonclassical moment of inertia I, is not an adequate model of an equilibrium superfluid. Unless abnormally small angular velocities are used (their size depending upon the radius of the bucket) the ideal Bose-Einstein gas would show equilibrium infrafluidity, not superfluidity.

However, the calculation is instructive in that it gives us a clue to the nature of a possible *nonequilibrium* superfluid state. Let us suppose that equilibrium is reached quickly for the highly excited states (excitation energy of order kT) but that the phase transitions discussed in Sec. 3, in which macroscopic numbers of particles shift from one very low-lying state to another, take a very long time to complete, a time much longer than the duration of usual experiments. Under these conditions, we get superfluid behavior of the ideal Bose-Einstein gas below its transition point. This nonequilibrium value of $\bar{L}$ as a function of ω is shown in Fig. 1 by the dashed line.

In this sense, therefore, the ideal Bose-Einstein gas can be used as a model of a superfluid: there is normal fluid behavior above the transition point, then a thermodynamic transition to a state in which superfluidity can show itself as a transport (nonequilibrium) phenomenon. Of course, the assumption of no interaction between the particles is not adequate to encompass transport phenomena; the concept of a mean free path has to be introduced, and the relaxation times for various kinds of momentum transfer must be studied in detail. Such a theory has been developed, following the suggestions made here, by Klemens.[13]

The results of this analysis of the ideal Bose-Einstein gas suggest strongly that the superfluid phenomena observed in actual liquid helium are also nonequilibrium phenomena; that the superfluid state is only metastable, not thermodynamically stable. The proof that this is indeed the case will be given in subsequent papers.

ACKNOWLEDGMENTS

We would like to thank Dr. M. R. Schafroth for innumerable, extremely helpful discussions, without which this work might not have been done at all. We are also indebted to Dr. P. G. Klemens for valuable discussions on estimates of relaxation times.

[13] P. G. Klemens (to be published). We are grateful to Dr. Klemens for many valuable discussions on estimates of relaxation times, in particular for calling to our attention the great importance of the rapid variation of relaxation times with the relative velocity v_n-v_s between normal fluid and superfluid. While the weakly interacting Bose-Einstein gas provides a model of superfluidity with many of the qualitative features correctly reproduced, we do not believe that this model is sufficient for actual liquid helium. The thermodynamic properties of liquid helium are so different from those of an ideal (or almost ideal) Bose-Einstein gas that we believe it is unlikely that transport properties, which generally depend upon details of the model much more than thermodynamic properties, are given correctly by the weakly interacting Bose-Einstein gas model.

Statistical Mechanics of Rotating Buckets*

J. M. Blatt, S. T. Butler, and M. R. Schafroth
The F. B. S. Falkiner Nuclear Research and Adolph Basser Computing Laboratories, School of Physics, The University of Sydney, Sydney, Australia
(Received April 4, 1955)

A system in temperature equilibrium is enclosed in a container and this container is set rotating. We show that the effective moment of inertia is the same as what one would compute in the usual way, for both classical and quantum statistical mechanics. This apparently trivial theorem has applications for the theories of superfluidity of liquid helium and of superconductivity of metals.

1. INTRODUCTION

NO one doubts that a bucket of water, set rotating in such a way that thermal equilibrium is maintained at all times (e.g., no turbulent currents are set up), would show the conventional relation between angular momentum L and angular velocity ω:

$$L=I\omega, \tag{1.1}$$

where I, the moment of inertia, has the classical value

$$I=I_0=NM\langle(x^2+y^2)\rangle_{\text{Av}}. \tag{1.2}$$

Here N is the number of molecules, M the mass of each molecule, x and y are coordinates measured from the axis of rotation (the z-axis), and the braces denote an average over the volume of the bucket occupied by the fluid.

Yet no one, to our knowledge, has proved this theorem under general conditions, and it is the purpose of this paper to do so. Contrary to one's first impression, the proof is by no means trivially easy, and the theorem has important applications for the theory of superfluidity of liquid helium and the theory of superconductivity. These applications will be discussed in subsequent papers.

In Sec. 2, we consider a statistical system set in linear motion; the method of discussing the statistical mechanics of such a system is developed in some detail; the fact that the apparent mass of the system is equal to the actual mass NM is a trivial consequence of Galilean invariance, but our discussion allows us to derive a nontrivial equipartition theorem for the kinetic energy associated with the bulk motion of the fluid. In Sec. 3, we develop the statistical mechanics of a statistical assembly inside a rotating container, and we define the moment of inertia. Section 4 contains a proof that this moment of inertia is equal to (1.2) in classical statistical mechanics. Before we go on to quantum statistical mechanics, we develop, in Sec. 5, the concept of a "correlation length," which turns out to be of crucial importance for all the subsequent work. Sections 6 and 7 contain a proof of (1.2) in quantum statistical mechanics; this proof is more restricted than the classical one, and exceptional systems (in particular, the ideal Bose-Einstein gas below its condensation point) exist which violate (1.2). Some discussion and conclusions are given in Sec. 8, but the important applications of this theorem are given in subsequent papers.

* Also supported by the Nuclear Research Foundation within the University of Sydney.

2. LINEAR MOTION AND THE EQUIPARTITION THEOREM

We consider a box of volume V filled with an assembly of N identical particles, each of mass M. We intend to develop the formalism of statistical mechanics appropriate to discuss linear motion of the whole system with constant linear velocity v. The fact that the linear momentum contributed by the fluid is NMv is of course a trivial consequence of Galilean invariance; but our formal approach will enable us to deduce a nontrivial equipartition theorem from this fact.

Let us first consider the statistical mechanics of the box *at rest*, paying special attention to the method of including the effect of the walls of the box. The Hamiltonian of the system is

$$H=\sum_{i=1}^{N}\mathbf{p}_i^2/2M+V(\mathbf{r}_1,\mathbf{r}_2,\cdots,\mathbf{r}_N). \tag{2.1}$$

The conventional procedure is the following: the potential energy V includes the effects of all forces between the particles of the fluid, but does *not* include the effects of forces between particles of the fluid and the walls of the box. These latter effects are taken into account approximately by replacing the actual, rough walls of the box by perfectly reflecting mirrors, or, in quantum-mechanical terms, by boundary conditions on the admissible wave functions. All wave functions must vanish whenever any one of the coordinates $\mathbf{r}_i$ lies on or outside the walls of the container. The free energy F of the system is then given by ($\beta=1/kT$)

$$\exp(-\beta F)=\text{Trace}[\exp(-\beta H)], \tag{2.2}$$

where the trace is to be taken over any complete set of wave functions which satisfy these particular boundary conditions. It is important to realize that the trace (2.2) is not even defined unless the boundary conditions on the wave functions are specified. Furthermore, in order to define thermodynamic quantities such as the free

energy which remain constant with time (in thermal equilibrium), it is necessary that the boundary conditions imposed on the wave functions be themselves time-independent.

Let us now return to the boundary conditions for the box at rest; what is the justification for replacing the actual rough walls of the box by perfectly reflecting mirrors? Consider a typical collision between a molecule of the fluid and the actual wall: the collision results in a reflection of the molecule back into the fluid, but this reflection is not specular. The transverse component of momentum is not preserved, because the wall is rough and can take up transverse momentum; similarly the normal component of the momentum is not merely changed in sign. Furthermore the wall itself is in heat motion, so that collision between a molecule and the wall will in general result in an exchange of energy as well as in an exchange of momentum. However, once thermal equilibrium has been established within the system the over-all, average effect of collisions between molecules and the wall is the same as if the wall were a perfectly reflecting mirror surface without heat motion. For any one wall collision in which a molecule loses energy, there is another wall collision in which some other molecule gains energy; for any one wall collision in which a molecule loses transverse momentum, another molecule gains transverse momentum by a wall collision, etc. Thus it is a good approximation, in thermal equilibrium, to replace the actual rough walls by perfectly reflecting mirror surfaces. On the other hand, this method is quite useless in considering the approach to thermal equilibrium after the wall of the box has been heated up slightly, to a temperature higher than the temperature of the fluid. For by replacing the wall by a perfect mirror, we have lost the very mechanism by which equilibrium is established in this case.

One may object: Why can one not dispense with all this, and simply include the coordinates and momenta of the molecules making up the wall into the Hamiltonian H? This procedure is contrary to the spirit of statistical mechanics, and is in principle impossible. The difficulty is merely removed one step, not overcome. The wall has an outside surface, and this outside surface interacts with its surroundings. If we do not wish to make approximations, the surroundings must now be included in the Hamiltonian as well, until we finally include the whole universe into the Hamiltonian. This is of course an impossible program. No matter how we do it, at a certain point we must make a cut between particles included in our system, and the rest of the world. This rest of the world *must* be treated schematically, in some approximation, not precisely by including it in the Hamiltonian. Once we realize that approximations of this kind are in principle necessary, there is no reason against making the simplest sensible approximation, namely the replacement of the walls by perfect mirrors.

Let us now consider the thermal equilibrium state of the fluid in a box moving with a constant linear velocity v in the x-direction. We again replace the walls of the box by perfectly reflecting mirrors, and write the effect of these mirror surfaces as boundary conditions on admissible wave functions. We observed before that, in order to do statistical mechanics, it is necessary that the boundary conditions imposed on the wave functions be independent of time. Since the boundary conditions refer to the walls of the container, the only way to fulfill this requirement is to use a set of coordinates in which the walls of the container correspond to time-independent values of the coordinates; i.e., in order to do statistical mechanics at all, we are *forced* to transform to a system of coordinates moving along with the container. This transformation is not one of convenience, which could be avoided, but is a necessary step in order to do statistical mechanics at all.

The transformation to the moving coordinate system is a canonical transformation; we obtain it most simply by writing down the Lagrangian corresponding to (2.1), and making the substitutions:

$$x_i' = x_i - vt, \quad y_i' = y_i, \quad z_i' = z_i. \tag{2.3}$$

Transforming back to the Hamiltonian formalism, we get the following expressions for the canonical momenta:

$$p_{xi}' = p_{xi}, \quad p_{yi}' = p_{yi}, \quad p_{zi}' = p_{zi}. \tag{2.4}$$

We note that the momentum variable p_{xi}' represents the linear momentum of particle number i with respect to the stationary coordinate system, *not* with respect to the moving coordinate system. The transformed Hamiltonian is

$$H' = H - v\sum_{i=1}^{N} p_{xi} = H - vP, \tag{2.5}$$

where P is the x-component of the total linear momentum of all particles. The Hamiltonian H' is *not* numerically equal to the original Hamiltonian H; this is a consequence of the fact that the transformation (2.3), (2.4) involves the time explicitly.

We now determine the linear momentum of the fluid in our moving box, in thermal equilibrium. This is given by

$$\bar{P}(v) = \frac{\mathrm{Trace}[P\exp(-\beta H')]}{\mathrm{Trace}[\exp(-\beta H')]} = \frac{\mathrm{Trace}\{P\exp[-\beta(H-vP)]\}}{\mathrm{Trace}\{\exp[-\beta(H-vP)]\}}. \tag{2.6}$$

The trace is to be taken over all wave functions of the desired symmetry (symmetric for Bose-Einstein statistics, antisymmetric for Fermi-Dirac statistics) which satisfy the boundary conditions at the walls of the box.

In classical statistical mechanics the trace represents an integral over the $6N$-dimensional phase space.

We evaluate (2.6) by making a second canonical transformation:

$$\begin{aligned} &x_i''=x_i', \qquad y_i''=y_i', \qquad z_i''=z_i', \\ &p_{xi}''=p_{xi}'-Mv, \quad p_{yi}''=p_{yi}', \quad p_{zi}''=p_{zi}'. \end{aligned} \tag{2.7}$$

The fact that it *is* a canonical transformation can be seen by writing down the commutators; the additional term Mv commutes with everything. When we substitute (2.7) into the effective Hamiltonian H', (2.5), we obtain the new Hamiltonian H'':

$$\begin{aligned} H'=H''&=\sum_{i=1}^{N}(\mathbf{p}_i'')^2/2M+V(\mathbf{r}_1'',\mathbf{r}_2'',\cdots,\mathbf{r}_N'')-\tfrac{1}{2}NMv^2 \\ &=H(p_i'',q_i'')-\tfrac{1}{2}NMv^2, \end{aligned} \tag{2.8}$$

where now $H(p_i'',q_i'')$ is formally the same function of its variables as the initial Hamiltonian H, (2.1), was of the variables p_i and q_i. This invariance of form is the expression of the Galilean invariance of the underlying physical problem. If we did not have Galilean invariance, it would be impossible to find a canonical transformation which brings the transformed Hamiltonian back to its original form [except for an irrelevant added constant in (2.8)]. The formulation in terms of the once-primed coordinates and momenta is definitely not Galilean invariant, since we had to choose one preferred (moving) system of coordinates. It will be essential that a canonical transformation akin to (2.7) is *not* possible for uniform *angular* motion of the container.

We now substitute (2.7) into the definition (2.5) of P, the total linear momentum, to get

$$P=NMv+\sum_{i=1}^{N}p_{xi}''=NMv+P'', \tag{2.9}$$

where P'' is the same function of the p_i'' as P was of the p_i. P'' is clearly the linear momentum of the fluid as measured by an observer at rest with respect to the box. Substituting (2.8) and (2.9) into (2.6) we get

$$\begin{aligned} \bar{P}(v)&=\frac{\exp(-\tfrac{1}{2}NMv^2)\,\mathrm{Trace}\{(P''+NMv)\times\exp[-\beta H(p'',q'')]\}}{\exp(-\tfrac{1}{2}NMv^2)\,\mathrm{Trace}\{\exp[-\beta H(p'',q'')]\}} \\ &=NMv, \end{aligned} \tag{2.10}$$

where the last result follows from the fact that the trace of P'' is zero as a result of the formal invariance: the trace of $P''\exp[-\beta H(p'',q'')]$ is formally identical with the trace of $P\exp[-\beta H(p,q)]$, and this latter trace is just the expectation value of the total linear momentum in a stationary box, which is zero.

So far, we seem to have done no more than to prove a well-known result by a long and tedious method. However, this trivial result has nontrivial consequences. Let us evaluate the "effective mass" of the fluid in the container, i.e., the derivative of $\bar{P}(v)$ with respect to v, in the limit $v\to 0$. According to (2.10) this derivative is equal to NM. But we can also evaluate the derivative directly from Eq. (2.6). The denominator of (2.6) does not contain a term linear in v, in spite of appearances. This can be seen by observing that motion to the left and motion to the right must be equivalent in their effect on the partition function (the denominator); or else it can be seen directly from (2.8): the additional term is of order v^2. Hence we get[1]

$$\left(\frac{d\bar{P}}{dv}\right)_{v=0}=\beta\frac{\mathrm{Trace}[P^2\exp(-\beta H)]}{\mathrm{Trace}[\exp(-\beta H)]}=\beta\langle P^2\rangle_{\mathrm{av};\ v=0}. \tag{2.11}$$

When we equate this value to the result NM obtained from (2.10), we get an equipartition theorem for the degree of freedom associated with bulk motion of the fluid:

$$\langle(P^2/2NM)\rangle_{\mathrm{av}}=\tfrac{1}{2}kT. \tag{2.12}$$

Hence we see that the degrees of freedom associated with bulk linear motion have the equipartition value of energy, even in quantum statistical mechanics. Yet we know that the equipartition theorem itself fails in quantum statistical mechanics. What, then, is the relationship between (2.12) and the equipartition theorem in general? To see this, we take the definition of P as the sum of all the p_{xi} and substitute it into (2.12). Since the particles are all identical, all square terms give the same result, and all cross terms are equal to each other; hence we get

$$\begin{aligned} \langle P^2/2NM\rangle_{\mathrm{av}}&=\langle p_{x1}^2/2M\rangle_{\mathrm{av}} \\ &\quad+(N-1)\langle p_{x1}p_{x2}/2M\rangle_{\mathrm{av}}=\tfrac{1}{2}kT. \end{aligned} \tag{2.13}$$

In classical statistical mechanics, there is no correlation between momenta of different particles. That is, $\langle p_{x1}p_{x2}\rangle_{\mathrm{av}}=0$ and the equipartition theorem is a consequence of (2.13). In quantum statistical mechanics, even an ideal gas of identical particles exhibits correlations between momenta of different particles, as a result of the Bose-Einstein (or Fermi-Dirac) statistics, i.e., as a result of the fact that the wave functions must satisfy special symmetry conditions. If we could take our trace over all wave functions, rather than over all symmetric (or antisymmetric) wave functions, the ideal (noninteracting) gas would obey the equipartition theorem even in quantum mechanics. The presence of interactions does *not* lead to correlations between momenta of different particles in classical statistical mechanics. In quantum statistical mechanics, the interparticle forces can produce momentum correlations,

[1] Strictly speaking, an error is made here, since the operator P does not commute with the Hamiltonian H in the presence of walls. A more detailed consideration shows that the error is of order $1/N$ (N=number of particles) and must therefore be ignored in statistical mechanics.

for example by producing bound states of pairs of particles which then form "molecules."

Equation (2.13) is the quantum-mechanical generalization of the equipartition theorem for a gas of identical particles. We can solve for the correlation function of momenta of different particles:

$$(N-1)\langle p_{x1}p_{x2}\rangle_{Av}=MkT-\langle p_{x1}^2\rangle_{Av}. \tag{2.14}$$

The deviations from classical equipartition are in opposite directions for Fermi-Dirac and Bose-Einstein statistics; the correlation coefficient $\langle p_{x1}p_{x2}\rangle_{Av}$ is negative for Fermi-Dirac statistics, positive for Bose-Einstein statistics.

A very interesting feature of (2.14) is the dependence of $\langle p_{x1}p_{x2}\rangle_{Av}$ on the number of particles N. We shall return to this point in some detail, in Sec. 5.

An equation similar to the virial theorem in classical statistical mechanics can also be derived, and again differs from the classical equation by the presence of correlation.

Finally, we can generalize to systems containing different kinds of particles. Let us take two kinds of particles, of masses M and M', respectively; there are N particles of the first kind, N' of the second kind. Then the generalization of (2.12) is

$$\langle P^2\rangle_{Av}=(NM+N'M')kT. \tag{2.15}$$

The corresponding generalization of (2.14) is not very useful, since it involves three different types of correlations.

3. DEFINITION OF THE MOMENT OF INERTIA

We now perform a similar analysis for rotational motion. Unlike the translational case, the result is not a consequence of Galilean invariance; the transformation to a rotating system of coordinates is not a Galilean transformation.

The same arguments which led us before to transform to a system of coordinates moving along with the box, now lead us to transform to a system of coordinates rotating along with the box. Just as before, the conditions of the problem *force* us to use this transformation, since otherwise we could not replace the actual walls of the container by perfectly reflecting mirror surfaces (boundary conditions on wave functions).[2]

The canonical transformation to rotating coordinates is similar to the transformation discussed in Sec. 2. We use cylindrical coordinates r, θ, z. The new coordinates are

$$\theta_i'=\theta_i-\omega t, \quad r_i'=r_i, \quad z_i'=z_i. \tag{3.1}$$

[2] It is interesting to observe that, unlike the translational motion of the box, there exists a special case here in which the roughness of the walls of the box is essential to the attainment of thermal equilibrium: consider a box with cylindrical symmetry, set in rotation about its symmetry axis. If the walls of the box are perfectly reflecting mirrors, the fluid inside the box is not set into rotation at all. However, even one small irregularity on the inside surface of the box is enough to make the thermal equilibrium state unique. This small surface irregularity takes the place of the "dust particle" in the theory of blackbody radiation.

Substitution into the Lagrangean corresponding to the Hamiltonian (2.1), and transformation back to the Hamiltonian formalism gives the transformed momenta

$$p_{\theta i}'=p_{\theta i}, \quad p_{ri}'=p_{ri}, \quad p_{zi}'=p_{zi}. \tag{3.2}$$

The canonical variable representing the transformed angular momentum is physically the angular momentum with respect to the rest frame, *not* with respect to the moving coordinate system. The transformed Hamiltonian is

$$H'=H-\omega\sum_{i=1}^{N}p_{\theta i}=H-\omega L, \tag{3.3}$$

where L is the total angular momentum of all particles around the z-axis. Again the transformed Hamiltonian is *not* numerically equal to the original Hamiltonian, nor is it equal to the Hamiltonian as written down by an observer moving along with the box.

It is instructive to write down this latter Hamiltonian. The observer moving along with the box would first of all introduce a term in the potential energy to describe the centrifugal force, i.e., he would write down the provisional Hamiltonian

$$H_1=(2M)^{-1}\sum_{i=1}^{N}\left(p_{ri}^2+p_{zi}^2+\frac{p_{\theta i}^2}{r_i^2}\right)+V-\tfrac{1}{2}M\omega^2\sum_{i=1}^{N}r_i^2. \tag{3.4}$$

This Hamiltonian fails to give the correct equations of motion because the Coriolis force is not included. The Coriolis force is a velocity-dependent force at right angles to the velocity vector, and is therefore entirely analogous to the magnetic force on a charged particle moving in a constant magnetic field (this is of course the basis of the Larmor theorem). Hence our moving observer would introduce a "vector potential" to describe the Coriolis force; the final Hamiltonian becomes

$$H''=(2M)^{-1}\sum_{i=1}^{N}\left[p_{ri}^2+p_{zi}^2+\frac{(p_{\theta i}-Mr_i^2\omega)^2}{r_i^2}\right]+V-\tfrac{1}{2}M\omega_i^2\sum_{i=1}^{N}r_i^2. \tag{3.5}$$

It is easily seen that this Hamiltonian does give the correct equations of motion. However, H'' is formally different from the Hamiltonian of the observer in the rest system, i.e., from H, Eq. (2.1). That is, H'' is not the same function of its variables as H is of its variables. The differences are twofold: there is the additional potential energy of the centrifugal force, and the alteration in the kinetic energy associated with the Coriolis force. The centrifugal force is proportional to ω^2 and can therefore be ignored for small values of ω. One might think that we could introduce the momentum

variables

$$p_{\theta i}'' = p_{\theta i} - Mr_i^2\omega, \tag{3.6}$$

which correspond to the angular momentum relative to the moving system, and thereby make the Hamiltonian H'' formally identical to H, (2.1), except for terms of order ω^2 (which are of no interest for the moment of inertia). However, unlike the corresponding Galilean transformation (2.7), (3.6) is *not* a canonical transformation. The new angular momenta introduced by (3.6) fail to commute with the radial momentum variables p_{ri}, and the commutators are of order ω, not ω^2.[3] This shows the essential difference between the rotational motion and the much simpler translational motion of the box.

The thermal equilibrium value of the angular momentum of the fluid in the rotating box is given by

$$\bar{L}(\omega) = \frac{\mathrm{Trace}[L\exp(-\beta H')]}{\mathrm{Trace}[\exp(-\beta H')]} = \frac{\mathrm{Trace}\{L\exp[-\beta(H-\omega L)]\}}{\mathrm{Trace}\{\exp[-\beta(H-\omega L)]\}}. \tag{3.7}$$

The denominator of (3.7) is the partition function of the rotating fluid, and is an even function of the angular velocity ω, by symmetry. We are interested in the moment of inertia, which is defined by

$$I = \left(\frac{d\bar{L}}{d\omega}\right)_{\omega=0}. \tag{3.8}$$

Unlike the translational case, where $\bar{P}$ is a linear function of v as long as v is small compared to the velocity of light, the angular momentum $\bar{L}$ is not a purely linear function of ω; as the angular velocity increases, centrifugal forces come into play, particles tend to be forced outward, and $\bar{L}$ increases more rapidly; that is, the effective moment of inertia is itself a function of ω. However, these effects are not involved in the definition (3.8) since we go to the limit $\omega \to 0$.

We now differentiate (3.7) with respect to ω, and then set ω equal to 0. The denominator is an even function of ω, and therefore its derivative with respect to ω is zero at $\omega = 0$. We then get the following definition of the moment of inertia of a fluid in thermal equilibrium:

$$I = \beta\frac{\mathrm{Trace}[L^2\exp(-\beta H)]}{\mathrm{Trace}[\exp(-\beta H)]}, \tag{3.9}$$

where H, Eq. (2.1), is the Hamiltonian of the fluid in a *stationary* container.

The following sections contain a proof, in classical and then quantum statistical mechanics, that I defined by (3.9) is equal to the conventional expression for the moment of inertia, Eq. (1.2).

[3] An expansion in powers of this commutator is possible, however, as has been shown by R. Peierls, Z. Physik **80**, 763 (1933).

4. THE MOMENT OF INERTIA IN CLASSICAL STATISTICAL MECHANICS

We first prove a lemma which is itself of some interest since it represents a considerable generalization of the equipartition theorem. To save writing, we introduce coordinates $q_1 = x_1$, $q_2 = y_1$, $\cdots$, $q_{3N} = z_N$, and we label the corresponding momentum components accordingly. Then the following lemma holds:

Lemma: Let $f(q_1, \cdots, q_{3N})$ be any function of the coordinates only. Then

$$\langle p_k p_{k'} f\rangle_{kk'} = \delta_{kk'} M_k kT\langle f\rangle_{kk'}. \tag{4.1}$$

The proof of the lemma is patterned along the conventional proof of the equipartition theorem. The average value on the left side of (4.1) is given by

$$\langle p_k p_{k'} f\rangle_{kk'} = \frac{\int d^{3N}q \int d^{3N}p\; p_k p_{k'} f\exp(-\beta H)}{\int d^{3N}q \int d^{3N}p\exp(-\beta H)}. \tag{4.2}$$

The Hamiltonian is a diagonal quadratic form in the momentum components, i.e.,

$$H = \sum_{k=1}^{3N}\frac{p_k^2}{2M_k} + V(q_1, \cdots, q_{3N}). \tag{4.3}$$

Hence we have the identity

$$p_k = M_k\partial H/\partial p_k. \tag{4.4}$$

Substitution into the numerator of (4.2) gives

$$\text{Numerator} = -(M_k/\beta)\int d^{3N}q \int d^{3N}p\; f p_{k'}\frac{\partial\exp(-\beta H)}{\partial p_k} = +\delta_{kk'}(M_k/\beta)\int d^{3N}q \int d^{3N}p\; f\exp(-\beta H), \tag{4.5}$$

where the second step comes from an integration by parts on p_k together with the observation that the function f is independent of p_k. Combination of (4.5) and (4.2) proves the lemma.

We are now in a position to prove the main theorem. We need the average value of L^2, and we therefore write

$$L^2 = \left(\sum_{i=1}^{N}(x_i p_{yi} - y_i p_{xi})\right)^2 = \sum_{i,j=1}^{N}(x_i p_{yi} - y_i p_{xi})(x_j p_{yj} - y_j p_{xj}). \tag{4.6}$$

Each term of (4.6) is of the form appropriate for use of our lemma (4.1). When we employ the lemma, all nondiagonal terms drop out identically, and we get

$$I=\beta\langle L^2\rangle_{\text{Av}}=\sum_{i=1}^{N} M_i\langle(x_i^2+y_i^2)\rangle_{\text{Av}} \tag{4.7}$$

which is what we set out to prove.

At first sight it may seem as if we have proved too much; for consider the following counter example: a cylindrical, closed container is filled partially with water and is mounted with its axis horizontal. The water settles in the lower part of the container. We now start rotating the container about its horizontal axis. The water stays near the bottom, and the observed angular momentum is not equal to $I\omega$ with I given by (4.7). The answer is that under these conditions thermal equilibrium is never reached. As the container rotates, the layer of water in contact with the wall moves along with the wall, and closed currents are set up in the water. These currents maintain themselves as long as the container is kept rotating. The viscous forces then produce heat in the water, and the temperature increases. Thus our theorem is not in contradiction to this case, simply because the theorem starts from the assumption of thermal equilibrium and is therefore not applicable to this case.

One condition of applicability of our theorem, i.e., the condition for the existence of a rotating motion of the fluid in thermal equilibrium, is that the potential energy V must be invariant under rotations.[4]

This is true, for example, for a rotating bucket of water with the axis of rotation parallel to the direction of the gravitational force, and is not true for a bucket of water mounted with the axis horizontal.

5. CORRELATION LENGTH

Unlike classical statistical mechanics, Eq. (4.7) does not hold exactly in quantum statistical mechanics. For normal systems, we shall show that (4.7) is approximately correct, the approximation becoming better and better as the size of the container is increased. However, there exists at least one exceptional system (the ideal Bose-Einstein gas below its condensation point) in which (4.7) is violated, no matter how big the size of the container.[5] It is clear, therefore, that some physical assumption about the fluid is necessary in order to prove (4.7) in quantum statistical mechanics. Thermal equilibrium alone is not enough.

We shall restrict ourselves to fluids (gases and liquids), and ignore solids. The moment of inertia of a solid is equal to the classical value immediately, as a result of the solidity, so we do not lose much in generality by restricting ourselves to fluids from the start. We point out, however, that our considerations *can* be applied to the electron gas in a metal, thereby enabling us to deduce some statements of interest for the theory of superconductivity (this will form the subject of a separate paper).

Let us consider two particles in our field, separated by a distance r_{12}. The correlation coefficient between their momentum components p_{x1} and p_{x2}, say, is defined by

$$Q_{xx}(r_{12})=\langle p_{x1}p_{x2}\rangle_{\text{Av}}/\langle\langle p_{x1}^2\rangle\rangle_{\text{Av}}^{\frac{1}{2}}\langle\langle p_{x2}^2\rangle\rangle_{\text{Av}}^{\frac{1}{2}}$$

with similar equations for Q_{xy}, etc. We assert that these correlation functions depend on the distance r_{12} between the two particles in the fluid, and in particular there must exist in any physical fluid a "correlation distance" Λ such that the correlations Q become negligibly small for $r_{12}\gg\Lambda$. Since this correlation distance forms the starting point of the entire subsequent discussion, we shall now devote some time to clarifying its meaning.

First of all, the correlations Q are identically zero in classical statistical mechanics. That is, two different particles of the fluid always have uncorrelated momentum components in thermal equilibrium, no matter how close together the particles are. Thus the correlation length Λ in classical statistical mechanics is zero. One might think *a priori* that momentum components of different particles should show a correlation if the particles are within one mean free path of each other, but the formalism of classical statistical mechanics shows that this is not so. The mean free path is a nonequilibrium concept, whereas the correlation length is an equilibrium quantity, and in classical statistical mechanics there is no relation between the two.

In quantum statistical mechanics, one may at first object to our definition of the correlation coefficients, on the grounds that the momenta and coordinates of the particles don't commute. However, it is still possible to define a quantity analogous to a probability of finding one particle at $\mathbf{r}_1$ with momentum $\hbar\mathbf{k}_1$ and another particle at $\mathbf{r}_2$ with momentum $\hbar\mathbf{k}_2$. This is done in the next section, Eq. (6.20). The analogy is imperfect, since the quantum mechanical expression is not only not positive definite, but even fails to be real. However, this does not affect the possibility of defining a correlation length, since the correlation coefficient Q can become small without being real or positive.

It was pointed out in Sec. 2 that there are two ways in which quantum mechanics leads to momentum correlations between particles of the fluid. First, there is the possibility of the formation of discrete bound states (molecules) in which the momenta of the atoms making up one molecule are strongly correlated. Second, the requirements of Bose-Einstein or Fermi-Dirac statistics entail momentum correlations even in an ideal

[4] There are two kinds of "rotational invariance": (1) Invariance of the form of the Hamiltonian under change of the origin of the angle variable θ, i.e., under the transformation $\theta'=\theta-\alpha$, and (2) Invariance of the form of the Hamiltonian under a transformation to rotating coordinates, $\theta'=\theta-\omega t$. The first invariance is the one we need here, and is frequently fulfilled in practice. The second invariance is never true. This differs from linear motion, where both translational invariance and Galilean invariance are observed.

[5] J. M. Blatt and S. T. Butler, preceding paper [Phys. Rev. **100**, 476 (1955)].

gas. Let us make a very rough first estimate of the correlation lengths we can expect from these two effects: the first effect gives rise to momentum correlations over distances of the order of magnitude of interatomic distances within molecules (we exclude solids from this discussion), the second effect can be estimated to give momentum correlations of significance over distances of the order of a de Broglie wavelength of a particle with energy $E=kT$.

These rough estimates are probably adequate for most systems, but they fail for some special systems. The most conspicuous failure occurs for the ideal Bose-Einstein gas below its condensation point. This system contains a macroscopic number of particles (comparable to the total number N) in a single-particle state of de Broglie wavelength comparable to the size of the box. Thus this de Broglie wavelength, rather than the de Broglie wavelength of a particle with energy $E=kT$, gives the most far-reaching momentum correlations, and momenta of different particles in the gas are correlated even when the particles are at opposite ends of the box, no matter how big we make the box. That is, the ideal Bose-Einstein gas below its condensation point does not have a finite correlation length (a correlation length independent of the size of the container).

The other ideal gases (Bose-Einstein gas above the condensation point, Fermi-Dirac gas) have in principle the possibility of infinite-range momentum correlations, because the same quantum state which is responsible for the long-range correlation effects in the condensed Bose gas exists in these other gases and can be occupied by particles of the gas. However, in the other ideal gases this quantum state, as well as all the other very long wavelength states, has such a small occupation probability that no significant correlation results from it.

When we turn from ideal gases (no interactions between the particles) to real physical fluids, one significant difference appears immediately: the interactions between the particles make correlations of *infinite* range (range as large as the container) impossible. For suppose the opposite were true; we could then take a volume of fluid as large as the interior of the sun, and would have to expect to find individual atoms with appreciably correlated momenta at opposite ends of this volume, in thermal equilibrium; if this be admitted as a possibility, let us make the container as large as the interior of a red giant star, or as large as the whole galaxy. It is clear that, no matter how weak the interactions between the particles are, they must eventually wash out correlation effects between very distant atoms of the fluid.

As a rough first approximation, applicable to dilute gases, we may assume that an upper limit to the correlation distance Λ is given by a few times the mean free path of a particle. That is, we assume that two particles which are many mean free paths away from each other have statistically uncorrelated momenta.[6] Of course, this upper limit may be much too high (as is shown by the example of classical statistical mechanics, where $\Lambda=0$ no matter how large the mean free path) and is not applicable to liquids, since the mean free path is not a useful concept in the liquid state. An upper limit to Λ for liquid helium will be established by special arguments in a subsequent paper. For the moment, we need not estimate the value of Λ; all we need is that a correlation distance Λ *exists* (i.e., is independent of the volume of the container) for any physical fluid.

There is then a significant qualitative difference between real fluids, no matter how weakly the particles interact with each other, and ideal, noninteracting gases. Real fluids always possess a correlation length Λ, and two particles in a real fluid are statistically uncorrelated when they are apart by a distance $r \gg \Lambda$; ideal gases *may* show long range correlations, and the ideal Bose-Einstein gas below its condensation point actually does so.[7]

It should be pointed out that the correlation length Λ is a quantity defined in the thermodynamic equilibrium state, not a transfer property. It is of course well known that the mean free path is of great importance in transfer phenomena; for example the flow of the fluid through a pipe of dimensions large compared to the mean free path is qualitatively different from the flow through a narrow pipe of diameter less than the mean free path (Knudsen regime). We are *not* concerned with this, but are pointing out qualitative differences between a "Knudsen gas" and a normal gas *in thermodynamic equilibrium.*

The ideal Bose-Einstein gas below its condensation point has a nonclassical moment of inertia[8]; we know from experience that most ordinary fluids have the classical moment of inertia. The discussion above has shown that there exists another property of fluids, the correlation length, which distinguishes the ideal Bose-Einstein gas below condensation from all other fluids. It is reasonable, therefore, to suppose that there is a connection between the moment of inertia and the correlation length. We shall show in the subsequent

[6] A single mean free path may not be enough, since significant momentum correlations might be maintained as the result of small chains of collisions. In principle, the correlation coefficient never becomes identically zero; however, the residual correlation coefficient at large distances decreases very rapidly with the distance (presumably exponentially) and therefore becomes negligible very quickly.

[7] The dangers of ignoring the infinite correlation length can be seen for example in the paper of P. R. Zilsel, Phys. Rev. **92**, 1106 (1953). Zilsel attempts to impose an external condition (prescribed value of the total linear momentum P) on a *part* of the volume of a container filled with an ideal Bose-Einstein gas below its condensation point. He then obtains results inconsistent with Galilean invariance (see Sec. 2 of this paper). The existence of a correlation length Λ in a real Bose-Einstein fluid implies important qualitative changes in the nature of the Bose-Einstein condensation phenomenon itself. However these qualitative changes are not of experimental importance provided the correlation length Λ is much larger than the mean distance between particles in the fluid.

mathematical sections that this is indeed the case. To be precise, let I be the actual moment of inertia defined by (3.9), and let I_0 be the classical moment of inertia, $\frac{1}{2}NMR^2$ for a cylindrical bucket of radius R. We shall show that[8]

$$I=I_0[1+\text{order}(\Lambda S/V)], \tag{5.1}$$

where S is the surface area of the container and V is its volume. If the fluid is normal, i.e., possesses a definite correlation length Λ, the correction term can be made arbitrarily small simply by increasing all linear dimensions of the container. But in the ideal Bose-Einstein gas below its condensation point, Λ is of the same order as the linear dimensions of the container, and the correction term remains important no matter how big we make the container.

In the later application, we shall need a somewhat stronger form of (5.1) for the special case that the container has the form of a cylinder of radius R and height D. We shall show in Appendix A that for such a cylinder the height D is unimportant, and hence

$$I=I_0[1+\text{order}(\Lambda/R)]. \tag{5.2}$$

The proof is relegated to an appendix because the formalism is simpler if *all* linear dimensions of the container, including the height, are large compared to the correlation length.

Having described our program, we continue the discussion of the correlation length. As soon as the particles in the fluid interact with each other, it is impossible to write the wave functions of the actual fluid as products of single-particle wave functions. Thus we can no longer talk of "occupation numbers" of single-particle states, and the free energy of the fluid cannot be written as a sum of contributions from single-particle states. However, it is still possible to define a quantity analogous to a probability of finding a particle at the point $\mathbf{r}$ with momentum equal to $\hbar\mathbf{k}$.[9]

It will be important in our subsequent mathematical work to use plane waves as our basic set of functions. Since the plane wave functions fail to satisfy the correct boundary conditions at the walls of the container, an error is made by this procedure, and this error is of the order of $\lambda_m S/V$, where λ_m is the largest de Broglie wavelength of importance in the partition function, S is the surface area of the container, and V is the volume of the container. This error estimate is the ratio of the volume in which the error is made, to the total volume. We shall now give arguments to prove that λ_m, the maximum de-Broglie wavelength of importance, is less than the correlation distance Λ.

We have mentioned two sources of momentum correlations in quantum statistical mechanics. One of these was the requirement of symmetry or antisymmetry of the wave functions. This requirement imposes momentum correlations of particles within one de Broglie wavelength of each other. Thus the correlation distance Λ must be at least as large as the maximum de Broglie wavelength of importance, i.e., $\lambda_m \leqslant \Lambda$. Of course, the correlation distance Λ may be much larger than this in special systems, but this does not concern us here.

Finally, we return to formula (2.14); since the average $\langle p_{x1}^2\rangle_{kl}$ is an intensive quantity the correlation coefficient $\langle p_{x1}p_{x2}\rangle_{kl}$ is seen to be inversely proportional to the number of particles in the system. In a fluid with correlation length Λ, this dependence is easily understood: when the two particles are farther apart than Λ, their momenta are uncorrelated, and the contribution to $\langle p_{x1}p_{x2}\rangle_{kl}$ vanishes. Thus the only regions of the two-particle configuration space $(\mathbf{r}_1,\mathbf{r}_2)$ which contribute significantly to $\langle p_{x1}p_{x2}\rangle_{kl}$ are such that $|\mathbf{r}_1-\mathbf{r}_2|<\Lambda$. The contributing volume is therefore Λ^3V rather than V^2, and the correlation coefficient $\langle p_{x1}p_{x2}\rangle_{kl}$ is proportional to Λ^3/V, i.e., inversely proportional to the volume and hence also inversely proportional to the number of particles N, in agreement with (2.14).

While (2.14) can be understood in terms of the correlation length, we emphasize that (2.14) holds even when no correlation length can be defined, for example, (2.14) holds for the ideal Bose-Einstein gas below its condensation point. The reason is that correlations, even over large distances, between *very small* momenta p_1 and p_2 make no effective contribution either to $\langle p_1^2\rangle_{kl}$ or to $\langle p_1p_2\rangle_{kl}$. Since long-range correlations are important only for states with long de Broglie wavelengths, and hence with small momenta, (2.14) does not depend for its validity upon the existence of a finite correlation length, and conversely the validity of (2.14) cannot be used to prove the existence of a finite correlation length for any particular fluid.

6. AVERAGES OF ONE-PARTICLE AND TWO-PARTICLE OPERATORS

In this section, we develop some general methods necessary to carry through the proof of (4.7) in quantum

[8] This theorem is correct, and will be proved to be so, on the assumption that the particles have no intrinsic spin (which is true for helium atoms). If the particles do have intrinsic spin, this spin may contribute to the total angular momentum of the rotating fluid, as a result of spin-orbit coupling. A more detailed consideration shows that this contribution is of order ω^2 for all systems except perhaps ferromagnets; a contribution of order ω^2 to the angular momentum is of no importance for the moment of inertia, Eq. (3.8). More important, however, is the fact that the angular momentum contributed by the spins is at most $N\hbar$, where N is the number of particles. (This value is obtained when all spins are lined up parallel to each other.) As the radius of the bucket is increased, the orbital contribution to the angular momentum is proportional to NR^2, hence for large enough buckets the spin contribution is always negligible compared to the orbital contribution. Since the spins are of no importance for the final result, we shall ignore spins from now on.

[9] This is the quantity $W(\mathbf{k}_1,\mathbf{r}_1)$ defined in Sec. 6. The analogy is not perfect because W can assume negative values. However, the expectation value of any operator F which is a sum of single-particle operators, $F=\sum_{i=1}^{N}F_i$, can be expressed in terms of W.

statistical mechanics. These methods are not without some interest apart from this particular problem.[10]

Let F be a sum of single-particle operators

$$F=\sum_{i=1}^{N} F_i, \tag{6.1}$$

where each F_i acts only on the coordinates of the ith particle. We are interested in the expectation value of F in thermal equilibrium:

$$\bar{F}=\frac{\text{Trace}[F\exp(-\beta H)]}{\text{Trace}[\exp(-\beta H)]} =N\frac{\text{Trace}[F_1\exp(-\beta H)]}{\text{Trace}[\exp(-\beta H)]}, \tag{6.2}$$

where the trace is to be taken over all symmetric wave functions for Bose-Einstein statistics, over all antisymmetric wave functions for Fermi-Dirac statistics.

We shall attempt to simplify the evaluation of this trace under the *assumption that there exists a correlation distance* Λ, of the kind discussed in Sec. 5. Let S be the surface area of the volume V in which our system is enclosed. We define the function $G(\mathbf{k},\mathbf{r})$ by

$$G(\mathbf{k},\mathbf{r})=\tfrac{1}{2}\exp(-i\mathbf{k}\cdot\mathbf{r})F_1\exp(+i\mathbf{k}\cdot\mathbf{r}) +\text{complex conj.} \tag{6.3}$$

We assert that there exists a real function $W(k^2)$ such that

$$\text{Trace}[F_1\exp(-\beta H)]=\sum_{\mathbf{k}_1}\int\frac{d^3r_1}{V}G(\mathbf{k}_1,\mathbf{r}_1)W(k_1^2) +\text{terms of order } \Lambda S/V. \tag{6.4}$$

The sum over $\mathbf{k}_1$ is a sum over plane wave states in the volume V; it can often be replaced by an integral over $\mathbf{k}_1$ space. The function W is the same for all operators F, and it is therefore in some sense a probability density for finding the wave vector $\mathbf{k}$ in an interval d^3k. However, W is not necessarily positive.

The proof of (6.4) proceeds in several stages. First of all, we have shown in Sec. 5 that the existence of a correlation distance allows us to perform the trace by summing over all plane wave functions.

Next we introduce notation. We let $\mathbf{R}$ stand for the ordered set of coordinate vectors

$$\mathbf{R}=(\mathbf{r}_1,\mathbf{r}_2,\cdots,\mathbf{r}_N), \tag{6.5}$$

and $\mathbf{K}$ for the ordered set of wave vectors

$$\mathbf{K}=(\mathbf{k}_1,\mathbf{k}_2,\cdots,\mathbf{k}_N), \tag{6.6}$$

and we use the short-hand notation

$$\exp(i\mathbf{K}\cdot\mathbf{R})=\exp[i(\mathbf{k}_1\cdot\mathbf{r}_1+\mathbf{k}_2\cdot\mathbf{r}_2+\cdots+\mathbf{k}_N\cdot\mathbf{r}_N)]. \tag{6.7}$$

Let Π be the permutation which replaces 1 by s_1, 2 by s_2, $\cdots$, N by s_N. Then the vector $\mathbf{R}^\Pi$ is defined by

$$\mathbf{R}^\Pi=(\mathbf{r}_{s1},\mathbf{r}_{s2},\cdots,\mathbf{r}_{sN}). \tag{6.8}$$

The permutation operator Π applied to the plane wave $\exp(i\mathbf{K}\cdot\mathbf{R})$ gives

$$\Pi\exp(i\mathbf{K}\cdot\mathbf{R})=\exp(i\mathbf{K}\cdot\mathbf{R}^\Pi). \tag{6.9}$$

We define ϵ_Π to equal $+1$ for Bose-Einstein statistics, to equal $(-1)^\Pi$ for Fermi-Dirac statistics. Furthermore, we define the symbol TRACE to mean a trace over all states, irrespective of symmetry; the symbol Trace refers to a trace over symmetric (Bose-Einstein statistics) or antisymmetric (Fermi-Dirac statistics) states only. We then get[11]

$$\begin{aligned}\text{Trace}[F_1\exp(-\beta H)] &=(N!)^{-1}\sum_\Pi \epsilon_\Pi\,\text{TRACE}[F_1\exp(-\beta H)\Pi]\\ &=\tfrac{1}{2}(N!)^{-1}\sum_\Pi \epsilon_\Pi\,\text{TRACE}[F_1\exp(-\beta H)\Pi\\ &\qquad +\Pi\exp(-\beta H)F_1],\end{aligned} \tag{6.10}$$

where the sum extends over all possible permutations Π, and the symmetrized form is preferable to exhibit manifest reality of the final quantities.

We now write down the trace in completely explicit form. There is a sum over all plane wave states, i.e., over all sets of vectors $\mathbf{k}_1$, $\mathbf{k}_2$, $\cdots$, $\mathbf{k}_N$. Furthermore, there is an integration over $3N$ dimensional configuration space to obtain the diagonal matrix elements for the trace. Thus we get, using the Hermitean properties of F_1, H, and Π

$$\begin{aligned}\text{Trace}[F_1\exp(-\beta H)]=\tfrac{1}{2}(N!)^{-1}\sum_\Pi \epsilon_\Pi\sum_{\mathbf{K}}\int\frac{d^{3N}r}{V^N}\\ \times\{[F_1\exp(i\mathbf{K}\cdot\mathbf{R})]^*\exp(-\beta H)\exp(i\mathbf{K}\cdot\mathbf{R}^\Pi)\\ +[\exp(-\beta H)\exp(i\mathbf{K}\cdot\mathbf{R}^\Pi)]^*F_1\exp(i\mathbf{K}\cdot\mathbf{R})\}.\end{aligned} \tag{6.11}$$

We define the function $G_1(\mathbf{k}_1,\mathbf{r}_1)$ by the relation

$$\begin{aligned}G_1(\mathbf{k}_1,\mathbf{r}_1)&=\exp(-i\mathbf{K}\cdot\mathbf{R})F_1\exp(i\mathbf{K}\cdot\mathbf{R})\\ &=\exp(-i\mathbf{k}_1\cdot\mathbf{r}_1)F_1\exp(i\mathbf{k}_1\cdot\mathbf{r}_1).\end{aligned} \tag{6.12}$$

The second form follows because the operator F_1 acts upon the coordinates of particle 1 only, by assumption. We then get

$$\begin{aligned}\text{Trace}[F_1\exp(-\beta H)]=\tfrac{1}{2}(N!)^{-1}\sum_\Pi \epsilon_\Pi\sum_{\mathbf{K}}\int\frac{d^{3N}r}{V^N}\\ \times[G_1^*(\mathbf{k}_1,\mathbf{r}_1)\exp(-i\mathbf{K}\cdot\mathbf{R})\exp(-\beta H)\\ \times\exp(i\mathbf{K}\cdot\mathbf{R}^\Pi)+\text{compl. conj.}].\end{aligned} \tag{6.13}$$

[10] Similar methods have been used elsewhere: E. P. Wigner, Phys. Rev. **44**, 31 (1931); J. E. Moyal, Proc. Cambridge Phil. Soc. **45**, 99 (1949); J. E. Mayer and W. Band, J. Chem. Phys. **15**, 141 (1947); H. S. Green, Proc. Roy. Soc. (London) **A194**, 244 (1948); J. Chem. Phys. **19**, 955 (1951). Our main interest here is in the symmetry properties of the functions W and U defined in this section, rather than in an explicit evaluation of these functions.

[11] B. Kahn and G. E. Uhlenbeck, Physica **5**, 399 (1938).

We now observe that the integrations over $\mathbf{r}_2$, $\mathbf{r}_3$, $\cdots$, $\mathbf{r}_N$ do not involve the function G_1 at all. The same holds true of the sums over $\mathbf{k}_2$, $\mathbf{k}_3$, $\cdots$, $\mathbf{k}_N$, and of the sum over permutations Π. Thus all these operations can be performed once and for all, independently of the form of G_1, i.e., independently of the operator F_1. We define a function $W(\mathbf{k}_1,\mathbf{r}_1)$ by

$$W(\mathbf{k}_1,\mathbf{r}_1)=(N!)^{-1}\sum_{\Pi}\epsilon_{\Pi}\sum_{\mathbf{k}_2\cdots\mathbf{k}_N}\int\frac{d^3r_2\cdots d^3r_N}{V^{N-1}}$$

$$\times\exp(-i\mathbf{K}\cdot\mathbf{R})\exp(-\beta H)\exp(+i\mathbf{K}\cdot\mathbf{R}^{\Pi}). \quad (6.14)$$

This is a definite function of $\mathbf{k}_1$ and $\mathbf{r}_1$, *not* a formal quantity which still involves operators acting on $\mathbf{r}_1$. For $\exp(-\beta H)\exp(i\mathbf{K}\cdot\mathbf{R}^{\Pi})$ is a definite function of the variables $\mathbf{r}_1$, $\mathbf{r}_2$, $\cdots$, $\mathbf{r}_N$, and this is multiplied by $\exp(-i\mathbf{K}\cdot\mathbf{R})$ and integrated over $\mathbf{r}_2$, $\cdots$, $\mathbf{r}_N$, thereby yielding a definite function of $\mathbf{r}_1$. $W(\mathbf{k}_1,\mathbf{r}_1)$ is very hard to evaluate explicitly; but we shall not need to do so in this paper.

We also observe that we have *not* replaced sums over wave vectors $\mathbf{k}$ by integrals over a continuous k-space. It is well known that this procedure leads to difficulties when the particles in the system can form bound states, and results in an incorrect omission of the contributions of the bound states to the trace. But as long as we keep to sums over discrete wave vectors $\mathbf{k}$, we are summing over a complete set of functions, and no states are omitted.

With this definition, we get

$$\mathrm{Trace}[F_1\exp(-\beta H)]=\tfrac{1}{2}\sum_{\mathbf{k}_1}\int\frac{d^3r_1}{V}[G_1^*(\mathbf{k}_1,\mathbf{r}_1)W(\mathbf{k}_1,\mathbf{r}_1)$$

$$+W^*(\mathbf{k}_1,\mathbf{r}_1)G_1(\mathbf{k}_1\mathbf{r}_1)]. \quad (6.15)$$

We now proceed to show that, to the approximation involved in neglecting terms of order $\Lambda S/V$, the function $W(\mathbf{k}_1,\mathbf{r}_1)$ is independent of $\mathbf{r}_1$, depends on $\mathbf{k}_1$ only through k_1^2, and is real.

To show this, we observe that the existence of the correlation distance Λ allows us, in the calculation of W, (6.14), to replace the true volume first of all by a cube of the same total volume, and then to replace the correct boundary conditions at the surfaces of the cube by the condition of periodicity of all wave functions. Both operations introduce errors of order $\Lambda S/V$, and are therefore allowed to us.

We now show that $W(\mathbf{k}_1,\mathbf{r}_1)$ is independent of $\mathbf{r}_1$ in the periodicity cube, provided only that the Hamiltonian operator H is invariant under all translations.[12] Let us consider the function $W(\mathbf{k}_1,\mathbf{r}_1+\mathbf{a})$ where $\mathbf{a}$ is a constant vector. In the definition (6.14) of W, we introduce new variables of integration: $\mathbf{r}_2'=\mathbf{r}_2-\mathbf{a}$, $\mathbf{r}_3'=\mathbf{r}_3-\mathbf{a}$, $\cdots$ $\mathbf{r}_N'=\mathbf{r}_N-\mathbf{a}$. Since we have imposed periodic boundary conditions, we do not need to change the limits of integration. The function $\exp(-i\mathbf{K}\cdot\mathbf{R})$ is then changed into

$$\exp\left[-i\sum_{j=1}^{N}\mathbf{k}_j\cdot\mathbf{a}\right]\exp(-i\mathbf{K}\cdot\mathbf{R}),$$

and the function $\exp(+i\mathbf{K}\cdot\mathbf{R}^{\Pi})$ is changed into

$$\exp\left[+i\sum_{j=1}^{N}\mathbf{k}_j\cdot\mathbf{a}\right]\exp(+i\mathbf{K}\cdot\mathbf{R}^{\Pi}).$$

The additional factors are just constants as far as the integrations over $\mathbf{r}_2$, $\mathbf{r}_3$, $\cdots$, $\mathbf{r}_N$ are concerned, and also as far as the operator $\exp(-\beta H)$ is concerned. Thus these factors cancel, and we conclude that $W(\mathbf{k}_1,\mathbf{r}_1)$ is independent of $\mathbf{r}_1$. Furthermore, W is a scalar function of the vector $\mathbf{k}_1$ only (it is independent of $\mathbf{r}_1$), and this means it must be a function of k_1^2 only.[13]

Finally, we observe that an additional integration of (6.14) over $\mathbf{r}_1$ makes the integral in (6.14) into an integration over the full $3N$-dimensional configuration space. We conclude from the Hermitian property of the operator $\exp(-\beta H)$ and of the operator $\sum_{\Pi}\epsilon_{\Pi}\Pi$ that

$$\int d^3x_1 W(\mathbf{k}_1,\mathbf{r}_1)=\int d^3x_1 W^*(\mathbf{k}_1,\mathbf{r}_1). \quad (6.16)$$

Since W is independent of $\mathbf{r}_1$, we conclude that W is a real function of k_1^2. This finishes the proof of the statements made in connection with Eq. (6.4).

We now turn to expectation values of operators involving the coordinates of *two* particles. Let F be an operator of type

$$F=\sum_{\substack{i,j=1\\(i\neq j)}}^{N}F_{ij}, \quad (6.17)$$

where F_{ij} involves the coordinates of particles i and j only. We have

$$\mathrm{Trace}[F\exp(-\beta H)]$$

$$=N(N-1)\,\mathrm{Trace}[F_{12}\exp(-\beta H)]. \quad (6.18)$$

[12] For our later use of these theorems, this condition is stronger than necessary, and is not always satisfied. For example, the Hamiltonian of water in a bucket under the action of gravity is not invariant under translations parallel to the force of gravity. We shall need only invariance under translations perpendicular to the axis of rotation of the bucket, and hence can allow $W(\mathbf{k}_1,\mathbf{r}_1)$ to depend upon z_1, but not upon x_1 or y_1. The modifications of the proof introduced by the possibility of these weaker conditions are rather trivial, and we shall indicate them in the Appendix.

[13] The reader may object on one of two grounds: (a) this is so trivial it needs no proof; (b) the proof is incorrect since the first step, i.e., the replacement of the actual volume by a cube of the same volume content, already assumed that W was independent of $\mathbf{r}_1$, i.e., we have assumed at the start what we tried to prove. Objection (a) can be countered by observing that it is not obvious *a priori* that $W(\mathbf{k}_1,\mathbf{r}_1)$ might not be complex with a phase factor $\exp(i\mathbf{k}_1\cdot\mathbf{r}_1)$; this has been excluded by our proof. Objection (b) is incorrect: the correlation distance Λ assures us that $W(\mathbf{k}_1,\mathbf{r}_1)$ is a "local" property, i.e., W depends only on conditions within a neighborhood of size Λ around the point $\mathbf{r}_1$; but a local property need not be independent of the locality. The replacement of the actual volume by a cube is possible because W is a local property, and does *not* presume that W is independent of position.

We define the function $G(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2)$ by an equation analogous to (6.12):

$$G(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2)=\exp(-i\mathbf{K}\cdot\mathbf{R})F_{12}\exp(i\mathbf{K}\cdot\mathbf{R})$$
$$=\exp[-i(\mathbf{k}_1\cdot\mathbf{r}_1+\mathbf{k}_2\cdot\mathbf{r}_2)]F_{12}$$
$$\times\exp[+i(\mathbf{k}_1\cdot\mathbf{r}_1+\mathbf{k}_2\cdot\mathbf{r}_2)], \quad (6.19)$$

and a function $U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2)$ by an equation analogous to (6.14):

$$U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2)$$
$$=(N!)^{-1}\sum_{\Pi}\epsilon_{\Pi}\sum_{\mathbf{k}_3,\mathbf{k}_4,\cdots,\mathbf{k}_N}\int\frac{d^3r_3d^3r_4\cdots d^3r_N}{V^{N-2}}$$
$$\times\exp(-i\mathbf{K}\cdot\mathbf{R})\exp(-\beta H)\exp(+i\mathbf{K}\cdot\mathbf{R}^{\Pi}). \quad (6.20)$$

Comparison of (6.14) and (6.20) gives the relation

$$W(\mathbf{k}_1,\mathbf{r}_1)=\sum_{\mathbf{k}_2}\int\frac{d^3x_2}{V}U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2). \quad (6.21)$$

By writing down the trace in (6.18) explicitly, we see that

$$\mathrm{Trace}[F_{12}\exp(-\beta H)]$$
$$=\tfrac{1}{2}\sum_{\mathbf{k}_1,\mathbf{k}_2}\int\frac{d^3r_1d^3r_2}{V^2}(G^*U+U^*G). \quad (6.22)$$

We now establish certain properties of the function U which we shall need in the next section. First of all, the definition of U, (6.20), together with the assumption of a correlation distance Λ independent of the volume of the container, allows us to assert that U is invariant under translations of $\mathbf{r}_1$ and $\mathbf{r}_2$ simultaneously. The detailed argument is exactly the same as for W and needs no repetition. Next, by our fundamental assumption of a correlation distance Λ, U becomes separable if the distance $|\mathbf{r}_1-\mathbf{r}_2|$ exceeds Λ. That is

$$U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2)=W(k_1^2)W(k_2^2) \quad \text{for} \quad |\mathbf{r}_1-\mathbf{r}_2|\gg\Lambda. \quad (6.23)$$

The fact that the factors are just $W(k_1^2)$ follows from separability together with Eq. (6.21). From the translation symmetry we conclude that

$$U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2)=U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_{12})$$
$$[\text{except near the boundaries}]. \quad (6.24)$$

When $\mathbf{r}_1$ and $\mathbf{r}_2$ are well separated, (6.23) applies and U is real since W is real. When $|\mathbf{r}_{12}|$ is of order Λ or smaller, U need not be real. However, U has additional symmetries as a result of being a "local" property: U is invariant under rotations of the set of vectors $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{r}_{12}$ simultaneously, and it is also invariant under the inversion operation:

$$U(-\mathbf{k}_1,-\mathbf{k}_2,-\mathbf{r}_{12})=U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_{12}). \quad (6.25)$$

This finishes our general analysis of expectation values of one- and two-particle operators. A word of caution is perhaps in order. The theorems may look more impressive than they really are. For most practical problems we need to know the temperature dependence of physical quantities in thermal equilibrium. This temperature dependence is contained in the functions W and U, and the symmetry properties of W and U give absolutely no clue to the temperature dependence of either. However, for our special purpose these symmetry properties are enough: we wish to prove a theorem true at all temperatures [Eq. (1.2)], and hence the proof need not involve the temperature explicitly.

7. MOMENT OF INERTIA IN QUANTUM STATISTICAL MECHANICS

Since the classical proof given in Sec. 4 is based upon a generalization of the equipartition theorem, and the equipartition theorem fails to hold in quantum statistical mechanics, our method of proof must be altered. We shall establish the following relation between the average value of L^2 and the mean square total linear momentum (in the x-direction, say):

$$\langle L^2\rangle_{Av}=\langle\langle(x_1^2+y_1^2)\rangle_{Av}\langle P_x^2\rangle_{Av}$$
$$\times[1+\text{terms of order}(\Lambda S/V)]. \quad (7.1)$$

The additional terms are not evaluated but tend to zero as the volume is made infinitely large in such a way that the ratio of surface area to volume becomes vanishingly small.[14] The proof depends upon the "correlation distance" Λ introduced in Sec. 5, and the surface terms are of relative order $\Lambda S/V$ where S is the surface area and V is the volume.

Once relation (7.1) is established, the result we want follows from (2.11), i.e., from the equipartition theorem for the center-of-gravity motion; this latter has been shown to hold in quantum statistical mechanics as well as in classical statistical mechanics.

Let us therefore proceed to prove (7.1). We write

$$l_i=p_{0i}/\hbar=-i(x_i\partial/\partial y_i-y_i\partial/\partial x_i), \quad (7.2)$$

and we use the permutation symmetry in order to write

$$\mathrm{Trace}(L^2e^{-\beta H})=N\hbar^2\,\mathrm{Trace}(l_1^2e^{-\beta H})$$
$$+N(N-1)\hbar^2\,\mathrm{Trace}(l_1l_2e^{-\beta H}). \quad (7.3)$$

The traces are understood to be over all symmetric wave functions for Bose-Einstein statistics, over all antisymmetric wave functions for Fermi-Dirac statistics. We now use the results of the preceding section. We have from (6.4):

$$\mathrm{Trace}[l_1^2\exp(-\beta H)]$$
$$=\sum_{\mathbf{k}_1}\int\frac{d^3r_1}{V}G(\mathbf{k}_1,\mathbf{r}_1)W(k_1^2)+\text{terms of order }\Lambda S/V$$
$$=\sum_{\mathbf{k}_1}\int\frac{d^3r_1}{V}(x_1^2k_{y1}^2+y_1^2k_{x1}^2-2x_1y_1k_{x1}k_{y1})W(k_1^2)$$
$$+\text{order }\Lambda S/V, \quad (7.4)$$

[14] We should note that this condition excludes the possibility of making a cylindrical volume infinitely large by increasing the height of the cylinder without increasing its radius.

where the second line follows from the definition (6.3) of the function G. We introduce the notations

$$\langle x^2\rangle_{Av}=V^{-1}\int x^2 d^3\mathbf{r},\quad \langle y^2\rangle_{Av}=V^{-1}\int y^2 d^3\mathbf{r},$$
$$\langle xy\rangle_{Av}=V^{-1}\int xy d^3\mathbf{r}. \tag{7.5}$$

It should be noted that these averages are *geometrical* averages, not averages over canonical ensembles. The geometrical averages arise because W is independent of position. With this notation we have

$$\mathrm{Trace}[l_1^2\exp(-\beta H)]=\sum_{\mathbf{k}_1}[\langle x^2\rangle_{Av}k_{y1}^2+\langle y^2\rangle_{Av}k_{x1}^2 -2\langle xy\rangle_{Av}k_{x1}k_{y1}]W(k_1^2). \tag{7.6}$$

We now use the fact that W depends on k^2 only:

$$\sum_{k_1} k_{x1}k_{y1}W(k_1^2)=0, \tag{7.7}$$

$$\sum_{k_1} k_{x1}^2W(k_1^2)=\sum_{k_1} k_{y1}^2W(k_1^2) =\hbar^{-2}\,\mathrm{Trace}[p_{x1}^2\exp(-\beta H)], \tag{7.8}$$

where the last part of Eq. (7.8) is true up to surface terms of order $\Lambda S/V$. We therefore get

$$N\hbar^2\,\mathrm{Trace}[l_1^2\exp(-\beta H)]=N\langle(x^2+y^2)\rangle_{Av} \times\mathrm{Trace}[p_{x1}^2\exp(-\beta H)]+\text{Order }\Lambda S/V. \tag{7.9}$$

If this were classical mechanics, the trace of p_{x1}^2 would be determined by the equipartition theorem, and this would be the end of the proof. However, in quantum statistical mechanics the cross term [the last term of Eq. (7.3)] does not vanish, and must be taken into account explicitly. To do so, we use Eq. (6.22) and the properties of the function $U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2)$ which were established in the preceding section. According to the definition of $G(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2)$, Eq. (6.19), we have for the operator $F_{12}=l_1l_2$:

$$G(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2)=x_1x_2k_{y1}k_{y2}+y_1y_2k_{x1}k_{x2} -x_1y_2k_{y1}k_{x2}-y_1x_2k_{x1}k_{y2}. \tag{7.10}$$

Since (7.10) is already real, Eq. (6.22) shows that only the real part of U contributes.

Consider now the first term of (7.10); its contribution is

$$\sum_{\mathbf{k}_1,\mathbf{k}_2}\int\int\frac{d^3\mathbf{r}_1d^3\mathbf{r}_2}{V^2}x_1x_2k_{y1}k_{y2}U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2)\equiv A. \tag{7.11}$$

In order to simplify this expression, we first use the separability property (6.23). Consider values of $\mathbf{r}_1$ and $\mathbf{r}_2$ such that $|\mathbf{r}_1-\mathbf{r}_2|>\Lambda$; we then perform the sum over $\mathbf{k}$ vectors, and get zero:

$$\sum_{\mathbf{k}_1,\mathbf{k}_2}k_{y1}k_{y2}U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2) \cong\Big(\sum_{\mathbf{k}_1}k_{y1}W(k_1^2)\Big)\Big(\sum_{\mathbf{k}_2}k_{y2}W(k_2^2)\Big)=0$$
$$\text{for}\quad|\mathbf{r}_1-\mathbf{r}_2|>\Lambda. \tag{7.12}$$

We now introduce, in (7.11), the difference vector $\mathbf{r}_{12}=\mathbf{r}_2-\mathbf{r}_1$ as a variable of integration. Because of (7.12), large values of $\mathbf{r}_{12}$ make no contribution, hence we can extend the integration over the variable $\mathbf{r}_{12}$ over a volume V centered at $\mathbf{r}_{12}=0$ (at $\mathbf{r}_1=\mathbf{r}_2$), rather than over the correct volume (which depends on $\mathbf{r}_1$); this procedure is correct except within a distance Λ of the walls, and hence the error made is of order $\Lambda S/V$. We thus write for the quantity A, (7.11):

$$A=\sum_{\mathbf{k}_1,\mathbf{k}_2}\int\frac{d^3\mathbf{r}_1}{V}\int\frac{d^3\mathbf{r}_{12}}{V} \times(x_1^2+x_1x_{12})k_{y1}k_{y2}U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_{12}). \tag{7.13}$$

We assert that the second term in the parenthesis gives zero upon integration. This follows from the inversion symmetry of U, Eq. (6.25); for consider the contribution made by $-\mathbf{k}_1$, $-\mathbf{k}_2$, $-\mathbf{r}_{12}$: the factor x_{12} changes sign, the factor $k_{y1}k_{y2}$ is unchanged, and so is U. Thus, in the integral over $\mathbf{r}_{12}$ and the sums over $\mathbf{k}_1$ and $\mathbf{k}_2$, we get pairwise cancellation of terms.

Thus we have

$$A=\sum_{\mathbf{k}_1,\mathbf{k}_2}\int\int\frac{d^3\mathbf{r}_1d^3\mathbf{r}_2}{V^2}x_1x_2k_{y1}k_{y2}U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_1,\mathbf{r}_2)$$
$$=\langle x^2\rangle_{Av}\sum_{\mathbf{k}_1,\mathbf{k}_2}\int\frac{d^3\mathbf{r}_{12}}{V}k_{y1}k_{y2}U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_{12})$$
$$=\langle x^2\rangle_{Av}\hbar^{-2}\,\mathrm{Trace}[p_{y1}p_{y2}\exp(-\beta H)]$$
$$=\langle x^2\rangle_{Av}\hbar^{-2}\,\mathrm{Trace}[p_{x1}p_{x2}\exp(-\beta H)], \tag{7.14}$$

where the last step is a result of rotational invariance.

A similar reduction carried out on the second term of (7.10) gives the same as (7.14) but with $\langle x^2\rangle_{Av}$ replaced by $\langle y^2\rangle_{Av}$.

Finally, the last two terms of (7.10) make no contribution. Consider the third term of (7.10). We first replace x_1y_2 by x_1y_1 and use the inversion symmetry of U to show that the correction vanishes. Next we use translation symmetry in order to write:

$$\sum_{\mathbf{k}_1,\mathbf{k}_2}\int\int\frac{d^3\mathbf{r}_1d^3\mathbf{r}_2}{V^2}x_1y_1k_{y1}k_{x2}U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_{12}) =\langle xy\rangle_{Av}\sum_{\mathbf{k}_1,\mathbf{k}_2}\int\frac{d^3\mathbf{r}_{12}}{V}k_{y1}k_{x2}U(\mathbf{k}_1,\mathbf{k}_2,\mathbf{r}_{12}).$$

To show that this is zero, we perform a simultaneous rotation of the vectors $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{r}_{12}$ around the x-direction, through 180 degrees. The factor k_{y1} changes sign, the factor k_{z2} is unchanged, and U is also unchanged. Since the operation "sum over $\mathbf{k}_1$, $\mathbf{k}_2$ and integrate over $\mathbf{r}_{12}$" is invariant under such a rotation of dummy variables, we conclude that the result vanishes.

Thus we finally have, up to terms of relative order $\Lambda S/V$,

$$N(N-1)\hbar^2 \operatorname{Trace}[l_1 l_2 \exp(-\beta H)] = N(N-1) \times \langle (x^2+y^2) \rangle_{\text{Av}} \operatorname{Trace}[p_{x1} p_{x2} \exp(-\beta H)]. \quad (7.15)$$

Combining (7.3), (7.9), and (7.15), we obtain the desired result (7.1). This concludes our proof.

It should be noted that the proof given here applies to systems composed of identical particles. The extension to systems containing particles of several types is trivial and we shall not give it here. The crucial condition is the existence of a correlation length Λ which is independent of the volume of the container. The identity of the particles and the type of statistics obeyed by the particles is unessential; assuming particles of one kind only simplifies the mechanics of writing down the proof, and this is why we made this assumption.

8. CONCLUSION

We have now shown that the moment of inertia of a system in thermal equilibrium is equal to the usually computed quantity, both in classical statistical mechanics and in quantum statistical mechanics. There is, however, a considerable difference between the conditions of applicability of the two proofs. The classical proof is applicable to an arbitrary system in thermal equilibrium, with the only proviso being that the rotating motion must allow thermal equilibrium also, i.e., that the potential energy V of the system (excluding wall forces) is invariant under rotations about the axis which is being used as an axis of rotation.

The quantum mechanical proof, on the other hand, depended crucially upon the existence of a "correlation distance" Λ between particles with correlated momentum vectors. The reader may feel that this restriction is not essential, and is merely necessary for our particular proof. This is not so: there exists a system (the ideal Bose-Einstein gas below its condensation point) for which the moment of inertia I, defined by (3.9), is *not* equal to the usual value (1.2).[6] The ideal Bose-Einstein gas below the condensation point is characterized by a correlation distance equal to the linear dimensions of the container. This is so because a macroscopic number of particles are condensed into a state with wavelength equal to the dimensions of the container. Thus not only have we proved the theorem about the moment of inertia for finite Λ, but we can also produce a counter-example to show that no similar proof is possible if Λ can become infinite (i.e., as large as the container).

There are peculiar systems other than the ideal Bose-Einstein gas in which the correlation distance Λ is comparable to the linear dimensions of the container. For example, consider a pipe in the shape of a torus, filled with water, and assume that the water is flowing steadily in a clockwise direction. Then the momenta of particles in different parts of the pipe are correlated by the drift motion, and Λ is as large as the linear dimensions of the torus. Such a system is excluded for our purposes by the fact that it is not in thermal equilibrium if the walls of the pipe are at rest. We recall that the definition (3.9) of the moment of inertia involves averages over the thermal equilibrium distribution in a container at rest.

However, there are also some equilibrium systems with large correlation distances. For example, consider a container partially filled with water, in the complete absence of a gravitational field. This is a two-phase system, consisting of water and water vapor. Let us assume that the walls of the container are perfectly reflecting, so that the water does not adhere to the walls. In thermal equilibrium we then have one very large ball of water, with volume proportional to the volume of the container,[18] and this ball of water executes Brownian motion inside the container. This Brownian motion has a kinetic energy equal to $3kT/2$, and thus becomes slower and slower as the size of the system is increased. Nevertheless, there is then a correlation between momenta of particles anywhere within the ball of water, and the radius of the ball is proportional to the linear dimensions of the container. It is clear, also, that this system has very peculiar properties when it is set into rotation. Already at very low rotational velocities, the water ball breaks up and a sheath of water forms next to the walls, as a consequence of the centrifugal force. We have neglected the centrifugal force throughout (since we have always restricted ourselves to terms of order ω, and have ignored terms of order ω^2), and thus our whole analysis becomes unphysical for this case.

This is a fairly general property of many-phase systems in which the different phases can be separated in space by the centrifugal force. However we are primarily interested in applying our theorem to liquid helium, which is a single-phase system in which such things cannot happen.

These exceptional systems show that one must be careful to establish, in any particular case, that the conditions of validity of our theorem are satisfied. That is, one must show that a finite correlation distance Λ, independent of the volume of the container, does indeed exist. We shall do so for liquid helium, and we shall discuss the consequences of this theorem for the theory

[18] It is understood, of course, that the number of particles N is increased whenever the volume is increased, in such a way that the number density N/V stays constant.

of liquid helium, in the following paper. Our theorem also has some interesting applications for the theory of superconductivity, and these will be discussed separately.

ACKNOWLEDGMENTS

We would like to express our appreciation to Professor H. Messel who brought us together as a team, and whose constant encouragement and never failing initiative created the stimulating atmosphere to which this work is due.

APPENDIX: THE ROTATING CYLINDER

So far we have assumed that all linear dimensions of the container are large compared to the correlation length Λ. In applying these considerations to liquid helium, we shall need to consider a situation in which the "bucket" is a very flat cylinder, of radius $R \gg \Lambda$ but height D of the same order as Λ. It is intuitively obvious that the height of the cylinder is of no importance in thermal equilibrium, but it seems preferable to show formally that the proof given in Secs. 6 and 7 can be modified so as to allow for arbitrary values of D, and then leads to the result

$$I = I_0[1+\text{order } (\Lambda/R)]. \tag{A.1}$$

The first modification is that we cannot use plane wave states of form $\exp[i(k_x x+k_y y+k_z z)]$, since the violation of the boundary conditions at $z=0$ and $z=D$ (the lower and upper plate of the cylinder, respectively) has serious consequences if $D \lesssim \Lambda$. However, we can use states of the form

$$\psi(\mathbf{k},\mathbf{r}) = \exp[i(k_x x+k_y y)] \sin(k_z z), \tag{A.2}$$

where $k_z = n\pi/D$, n integral. These functions obey the boundary conditions at $z=0$ and $z=D$, but violate the boundary condition at $r=R$. This latter violation is not serious if $R \gg \Lambda$, and leads to errors of the order of Λ/R.

The (unsymmetrized) product wave functions for the N-particle system are defined by

$$\Psi(\mathbf{K},\mathbf{R}) = \prod_{i=1}^{N} \psi(\mathbf{k}_i,\mathbf{r}_i), \tag{A.3}$$

where $\mathbf{K}$ is the ordered ensemble of the N vectors $\mathbf{k}_i$, and $\mathbf{R}$ the ordered ensemble of the N vectors $\mathbf{r}_i$, as before. In order to continue, we shall need notation for the projections of all these vectors on the x-y plane. We shall use Gothic letters for this purpose, e.g., the symbol $\mathfrak{r}_1$ is the projection on the x-y plane of the three-dimensional vector $\mathbf{r}_1$; i.e., $\mathfrak{r}_1$ has components $(x_1,y_1,0)$. We then write (A.3) in the form

$$\Psi(\mathbf{K},\mathbf{R}) = \exp(i\mathfrak{K}\cdot\mathfrak{R}) \prod_{i=1}^{N} \sin(k_{zi} z_i). \tag{A.4}$$

We arrive at Eq. (6.11) just as before, but with $\exp(i\mathbf{K}\cdot\mathbf{R})$ replaced by $\Psi(\mathbf{K},\mathbf{R})$. At this stage, we restrict outselves to operators F_1 which do not act on the coordinate z_1 at all. We can therefore write

$$\begin{aligned} F_1\Psi(\mathbf{K},\mathbf{R}) &= \prod_{i=1}^{N} \sin(k_{zi} z_i) F_1 \exp(i\mathfrak{K}\cdot\mathfrak{R}) \\ &= \Psi(\mathbf{K},\mathbf{R}) \exp(-i\mathfrak{k}_1\cdot\mathfrak{r}_1) F_1 \exp(i\mathfrak{k}_1\cdot\mathfrak{r}_1) \\ &= \Psi(\mathbf{K},\mathbf{R}) G_1(\mathfrak{k}_1,\mathfrak{r}_1). \end{aligned} \tag{A.5}$$

The function G_1 introduced by (A.5) is the analog of the three-dimensional $G_1(\mathbf{k}_1,\mathbf{r}_1)$ defined by (6.12). With this definition of G_1, (6.13) is again correct (with the obvious replacement of $\exp(i\mathbf{K}\cdot\mathbf{R})$ by $\Psi(\mathbf{K},\mathbf{R})$ throughout). But in order to get a useful form, we now perform not only all the sums and integrations indicated in (6.14), but also a sum over k_{z1} and an integral over z_1. The two-dimensional analog of the function W, (6.14) is

$$W(\mathfrak{k}_1,\mathfrak{r}_1) = (N!)^{-1} \sum_{\Pi} \epsilon_\Pi \sum_{k_{z1}} \sum_{\mathbf{k}_\Pi,\cdots,\mathbf{k}_N} \int \frac{dz_1}{D} \int \frac{d^2\mathfrak{r}_2 \cdots d^2\mathfrak{r}_N}{V^{N-1}} \times \Psi^*(\mathbf{K},\mathbf{R}) \exp(-\beta H) \Psi(\mathbf{K},\mathbf{R}^\Pi) \tag{A.6}$$

(6.15) is then replaced by

$$\begin{aligned} &\text{Trace}[F_1 \exp(-\beta H)] \\ &= \tfrac{1}{2} \sum_{\mathfrak{k}_1} \int \frac{d^2\mathfrak{r}_1}{\pi R^2} [G_1^*(\mathfrak{k}_1,\mathfrak{r}_1) W(\mathfrak{k}_1,\mathfrak{r}_1) \\ &\qquad + G_1(\mathfrak{k}_1,\mathfrak{r}_1) W^*(\mathfrak{k}_1,\mathfrak{r}_1)]. \end{aligned} \tag{A.7}$$

The subsequent proof that $W(\mathfrak{k}_1,\mathfrak{r}_1)$ is invariant under translations of the two-dimensional vector $\mathfrak{r}_1$, depends on $\mathfrak{k}_1$ only through $\mathfrak{k}_1^2 = k_{x1}^2 + k_{y1}^2$, and is real, goes through without difficulty. The function $G(\mathfrak{k}_1,\mathfrak{r}_1)$ is then defined to be the real part of $G_1(\mathfrak{k}_1,\mathfrak{r}_1)$, and Eq. (6.4) is replaced by

$$\begin{aligned} &\text{Trace}[F_1 \exp(-\beta H)] \\ &= \sum_{\mathfrak{k}_1} \int \frac{d^2\mathfrak{r}_1}{\pi R^2} G(\mathfrak{k}_1,\mathfrak{r}_1) W(\mathfrak{k}_1^2)[1+\text{order}(\Lambda/R)]. \end{aligned} \tag{A.8}$$

The modification of the analysis of expectation values of two particle operators is completely similar. We replace vectors by their x-y plane projections systematically, and integrate over z_1, z_2 and sum over k_{z1}, k_{z2} before defining the two-dimensional analog of the function U, Eq. (6.20). The separability (6.23), translational invariance (6.24), and inversion invariance (6.25) have direct two-dimensional analogs. There is also rotation invariance under joint rotations of $\mathfrak{k}_1$, $\mathfrak{k}_2$, and $\mathfrak{r}_{12}$ about the z-direction. However, one invariance property is no longer possible: we cannot make a rotation by 180 degrees around the x-axis. Such a

rotation leaves x-components invariant, and changes y- and z-components into their negatives. Since z-components are involved, there is no direct two-dimensional analog. However, we did not really need the change of sign of z-components for the proof. The two-dimensional symmetry operation which is needed is a mirroring in the x-z plane, which leaves x-components invariant, and changes y-components into their negatives. The two-dimensional U-function, $U(\mathbf{f}_1,\mathbf{f}_3,\mathbf{r}_{12})$, is invariant under this mirror operation.

From hereon the proof given in Sec. 7 goes through without change (except for the replacement of the 180 degree rotation about x by the mirroring operation). We thus obtain the desired result (A.1), in which the height of the cylinder does not enter into the estimate of error.

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Nonequilibrium Nature of the Superfluid State

S. T. Butler and J. M. Blatt
The F. B. S. Falkiner Nuclear Research and Adolph Basser Computing Laboratories, School of Physics, The University of Sydney, Sydney, Australia*

(Received April 4, 1955)

We show that the result of the preceding paper is directly applicable to liquid helium, i.e., no superfluid state can exist in liquid helium under conditions of thermodynamic equilibrium. The observed superfluid properties of liquid helium must therefore be due to a nonequilibrium (metastable) superfluid state. Previous theoretical arguments for equilibrium superfluidity are examined, and an experimental test of nonequilibrium superfluidity is suggested.

1. INTRODUCTION

LIQUID helium below the λ point has superfluid properties, and many widely different theories have been advanced to account for this behavior. There is practically universal agreement in the theoretical literature, however, that the superfluid state of liquid helium is a state of thermal equilibrium.[1] The position is stated very clearly by Zilsel[2]: "The superflow is truly reversible; it is maintained not by the absence of collisions, but—under the given conditions—is the thermodynamic equilibrium state established by the collisions." Or, as Landau[3] puts it: "It is most essential that there is no friction between these two liquids moving through each other. . . . We get this relative motion when considering the statistical equilibrium in a uniformly rotating vessel. But if there can be some sort of relative motion in the state of statistical equilibrium, it means that it cannot be accompanied by friction." London[4] remarks that this point is still open to question as far as experimental confirmation is concerned, but considers the evidence for truly reversible (i.e., equilibrium) superflow extremely strong.

We have shown[5] that the ideal Bose-Einstein gas below its transition point does indeed have a peculiar relationship between the thermal equilibrium value $\bar{L}(\omega)$ of the angular momentum and the angular velocity ω of the container. In particular, the moment of inertia I, defined by

$$I=\lim_{\omega\to 0}(\partial\bar{L}/\partial\omega) \tag{1.1}$$

is less than the classical value, in spite of the fact that the system is in thermal equilibrium. However, a more detailed discussion showed[5] that the system would appear to be "infrafluid" in actual experiments, because the experiments average over rather wide ranges of the angular velocity ω. It was proposed that the weakly interacting Bose-Einstein gas might serve as a model of a *nonequilibrium* superfluid, that is, under certain reasonable assumptions about collision cross sections, the superfluid state would be metastable with a lifetime longer than the times used in the relevant experiments. This idea has been elaborated since by Klemens.[6]

The investigation of the behavior of the ideal Bose-Einstein gas suggested strongly that the superfluid state of liquid helium is not an equilibrium state, that it is maintained by the absence (or rather, the scarcity) of collisions, rather than being the thermodynamic equilibrium state established by the collisions. How-

* Also supported by the Nuclear Research Foundation within the University of Sydney.

[1] The nearest we could find to the contrary is in a paper by N. F. Mott, Phil. Mag. **40**, 61 (1949). Mott first reviews arguments for an energy gap between the ground state and all excited states of an interacting Bose-Einstein gas, and then goes on to discuss superfluidity in terms of a metastable state. However, if there really were an energy gap, the superfluid state would be thermodynamically stable. Thus Mott's discussion of the rotating toroid is inconsistent with his earlier arguments for an energy gap.

[2] P. R. Zilsel, Phys. Rev. **92**, 1106 (1953).

[3] L. Landau, J. Phys. (U.S.S.R.) **5**, 71 (1941).

[4] F. London, *Superfluids* (John Wiley and Sons, Inc., New York, 1954), Vol. 2.

[5] J. M. Blatt and S. T. Butler, this issue [Phys. Rev. **100**, 476 (1955)]. See also Phys. Rev. **96**, 1149 (1954).

[6] P. G. Klemens (to be published).

ever, this is no proof for actual liquid helium, which is very far from being an ideal Bose-Einstein gas. After some attempts to estimate relaxation times for actual actual helium,[7] we decided upon the opposite course of proving that the equilibrium state must have a normal moment of inertia. Assuming this has been proved, the logic is: since the equilibrium state has the normal moment of inertia, and since liquid helium shows an abnormal (superfluid) moment of inertia, the superfluid state of liquid helium cannot be an equilibrium state. The preceding paper[8] contains a general proof that a physical fluid inside a rotating bucket of radius R in thermal equilibrium has the classical relation between angular momentum and angular velocity, provided that R is very much larger than a length characteristic of the fluid, the so-called "correlation length" Λ. The abnormal moment of inertia of the ideal Bose-Einstein gas is not in contradiction to this theorem, but merely indicates that the ideal Bose-Einstein gas is an unphysical abstraction, with an infinite correlation length.

This paper contains the application of the rotating bucket theorem to liquid helium; first we derive an upper limit for the correlation length Λ in liquid helium from experiments on second sound; we next show that some experiments actually performed on liquid helium are essentially equivalent to measuring the angular momentum of the liquid inside a rotating bucket; finally the upper limit on Λ is less than the dimensions used in these experiments. We can therefore apply the rotating bucket theorem, and conclude that the superfluid state of liquid helium is not the thermodynamic equilibrium state, but only a metastable state. If one waits for thermodynamic equilibrium to establish itself, the superfluid properties of helium disappear; in particular, the two-fluid theory[4,9] of liquid helium is *not* applicable to the state of thermal equilibrium, but describes a metastable state.

Our proof that the superfluid state is metastable is not a complete theory of superfluidity; but it is the first necessary step in such a theory, since it shows what one must look for. The second step is a thorough understanding, starting from first principles, of the thermodynamic (equilibrium) properties of liquid helium. This second step has been carried through successfully by us, and will be published soon. The third step, i.e., the construction of a kinetic theory of the superfluid properties of liquid helium,[10] is likely to take considerable time, since at present no general theory of nonequilibrium processes exists which could be applied directly to a system as complicated as actual liquid helium.

2. CORRELATION LENGTH OF LIQUID HELIUM

In his fundamental paper, Landau[3] was the first to discuss the thermal equilibrium of liquid helium inside a rotating container. He predicted that the angular momentum $\bar{L}$, though proportional to the angular velocity ω, would be less than expected classically. The derivative $\partial\bar{L}/\partial\omega$ is the moment of inertia of the fluid. If the fluid splits into two components, a "superfluid" which remains stationary and a "normal fluid" which rotates along with the container, the ratio I/I_0 between the observed moment of inertia and the classically expected value is equal to the concentration of normal fluid:

$$\rho_n/\rho = I/I_0, \tag{2.1}$$

where ρ_n is the density of the normal fluid component, and ρ is the total density of liquid helium. Equation (2.1) may be considered the *definition* of ρ_n/ρ.

We have shown[8] that the thermal equilibrium moment of inertia I of any physical fluid in a cylinder of radius R is given by

$$I = I_0[1+\text{order } (\Lambda/R)], \tag{2.2}$$

where I_0 is the classical moment of inertia, and Λ is a length characteristic of the fluid, the "correlation length."

The correlation length Λ is defined in terms of the correlation between momenta $\mathbf{p}_1$ and $\mathbf{p}_2$ of two different atoms of the fluid, as a function of their distance r_{12}. Two atoms which are close together have highly correlated momenta, the correlation being due, for example, to the possibility of collisions between them, or of chains of collisions in which our two atoms are the first and last links, respectively. On the other hand, when two atoms are separated by a sufficiently large distance, we expect that their momenta $\mathbf{p}_1$ and $\mathbf{p}_2$ are essentially uncorrelated, i.e., the correlation function decreases rapidly once r_{12} exceeds some minimum distance Λ. This distance is the correlation length.

One may be tempted to argue from first principles

[7] We would like to thank Dr. Klemens for valuable discussions on estimates of relaxation times, which convinced us that it is very hard indeed to make reliable estimates without a very detailed knowledge and understanding of the thermodynamic and kinetic properties of the particular substance.

[8] Blatt, Butler, and Schafroth, preceding paper [Phys. Rev. **100**, 481 (1955)].

[9] L. Tisza, Phys. Rev. **72**, 838 (1947); L. Landau, J. Phys. (U. S. S. R.) **8**, 1 (1944), and many other papers.

[10] It is, in our opinion, impossible to make a valid theory of kinetic properties until the equilibrium properties are well understood. The partial agreement between the model of a weakly interacting Bose-Einstein gas and actual liquid helium found by Klemens, reference 6, was thus probably somewhat accidental. The mean free path needed by Klemens is of the order of 10^{-8} cm, which is less than the average spacing between helium atoms. Thus the approximation of weak interactions is inconsistent in this model. We believe that a more detailed theory of the kinetic properties of liquid helium may well turn out to reduce to an equivalent weakly interacting Bose-Einstein gas; however, it is very likely that the equivalent gas has a different relation between energy and momentum than $E=p^2/2M$ and has peculiar collision cross sections (which, for example, might depend on the energy of the relative motion between the particle considered and the mean motion of the fluid, rather than merely on the relative velocity of the two colliding particles). Until these matters are better understood, it is hard to say to what extent the partial agreement between the simpleminded weakly interacting Bose-Einstein gas model[6] and experiment is significant.

that atoms of helium are bound to collide with each other so frequently, and with such a short mean free path, that the correlation length is necessarily of the order of interatomic distance, i.e., of the order of 10^{-7} cm. Such an argument, while presumably correct for ordinary liquids, ignores the fact that liquid helium is known to be peculiar in other ways; hence the argument fails to carry conviction, and we even believe it to be wrong objectively; i.e., we suspect that the actual correlation length in liquid helium is appreciably larger than the mean interatomic distance.[11] The source of this long-range correlation is presumably connected with the Bose-Einstein statistics.

Rather, we must base our argument upon some property characteristic of superfluid liquid helium. We shall choose second sound for this purpose. As is well known[9,12] second sound is a "temperature wave" in which the total density ρ of the fluid does not vary appreciably from point to point. The densities ρ_n of the normal and ρ_s of the superfluid vary periodically in such a way that the total density

$$\rho=\rho_n+\rho_s \tag{2.3}$$

remains constant. Furthermore, with each fluid component is associated a velocity field, $\mathbf{v}_n$ and $\mathbf{v}_s$ for the normal and superfluid, respectively. These velocities also vary periodically in space and time, in such a way that the net mass motion vanishes:

$$\rho_n\mathbf{v}_n+\rho_s\mathbf{v}_s=0. \tag{2.4}$$

Second sound was predicted by Landau[9] and Tisza[13] independently; the observed properties of second sound (sound velocity as function of the temperature, thermal Rayleigh disk experiments, etc.[12]) leave no doubt that the mathematical expressions used by the two-fluid theory represent the data quite accurately, that the division into two fluids is reasonable, and that in particular Eqs. (2.3) and (2.4) hold in a good approximation. Of course, the experiments on second sound cannot decide whether the division into normal and superfluid flow is possible in complete thermodynamic equilibrium, or only in a state of "hindered" equilibrium, i.e., a metastable state of lifetime long compared to the period of the second sound.

It may appear doubtful to use a nonequilibrium phenomenon, such as second sound, in order to put an upper limit on an equilibrium quantity such as the correlation length Λ. However, it is well known that the usual two-fluid theory of liquid helium describes second sound motion in terms of thermodynamic quantities (such as the entropy and the specific heat of the liquid) and arrives at an experimentally correct relation for the velocity of second sound. That is, the usual two-fluid theory assumes that thermodynamic equilibrium is established within volumes small compared to λ^3, where λ is the wavelength of second sound. A similar assumption is commonly made for the description of ordinary sound as well. Since the two-fluid theory works well for actual second sound, we can use second sound to infer properties of the state of thermodynamic equilibrium. This is true even if the local equilibrium established in second sound motion is not a full thermodynamic equilibrium, but rather a "hindered" equilibrium.

Equation (2.4) shows that the superfluid component is not merely a passive substrate in second sound, but plays an active role. The equation makes sense only if it is possible to define a *local* superfluid velocity v_s, which varies in space and time so as to satisfy (2.4). "Local" here means over volumes of the fluid small compared to λ^3, where λ is the wavelength of second sound.[14] Since the superfluid component is actively involved in second sound propagation, an estimate of the correlation length Λ based on second sound is permissible.

We now use the observed properties of second sound, in particular the absence of dispersion in second sound propagation, to put an upper limit on the correlation length. We shall establish that dispersion of second sound (i.e., a nonlinear relation between frequency and wavelength) must be expected as soon as the wavelength of second sound becomes smaller than the correlation length Λ. For particles which are closer together than Λ have strongly correlated momenta, so that a change in the mean momentum of one of them automatically entails a change in the mean momentum of the second one. This statistical coupling between separated particles makes the equations for sound propagation nonlocal (integral rather than differential equations) and dependent on detailed statistical correlations (rather than merely local thermodynamic quantities) as soon as the wavelength becomes comparable to, or smaller than, the correlation length. Such effects are well-known in ordinary sound propagation when the wavelength of the sound wave becomes comparable to the mean free path of atoms in the gas.[15] For wavelengths much smaller than the correlation length the statistical correlations tend to "stiffen" the superfluid, and hence lead to strong dispersion of second sound. No such effects have been observed at temperatures above 0.4°K. We therefore conclude that the correlation length Λ in

[11] Otherwise it would be difficult to understand why the sharp thermodynamic phase transition of the ideal Bose-Einstein gas is not washed out altogether in actual liquid helium; experimentally, the phase transition at the λ point is perfectly sharp.

[12] J. G. Daunt and R. S. Smith, Revs. Modern Phys. **26**, 172 (1954).

[13] L. Tisza, Compt. rend. **206**, 1035, 1186 (1938); J. phys. radium **1**, 165, 350 (1940).

[14] This incidentally shows immediately that the extreme ideal Bose-Einstein gas picture is not a possible model for actual liquid helium; in the ideal Bose-Einstein gas model the superfluid component is identified with the condensed phase, i.e., with the particles in the ground state. The ground state wave function has a de Broglie wavelength comparable to the linear dimensions of the container, so that a *local* superfluid density or superfluid velocity does not exist.

[15] L. Bergmann, *Ultrasonics* (G. Bell and Sons, Ltd., London; 1938).

liquid helium is appreciably smaller than the smallest wavelength of second sound used in the experiments.[12] This upper limit on Λ is at present about 10^{-2} cm. It would be highly interesting to push the measurements to higher frequencies; presumably dispersion of second sound should become important at wavelengths of the order of 10^{-4} to 10^{-5} cm, which would be a reasonable order of magnitude for the correlation length.

3. PROOF OF NONEQUILIBRIUM SUPERFLUIDITY

We recall that Landau[3] predicted an abnormally low value of the moment of inertia of liquid helium in a rotating bucket. After the war, Andronikashvili[16] set out to test Landau's prediction experimentally. Andronikashvili's experimental setup was designed to be as close as practicable to Landau's rotating bucket. It consisted of a pile of closely spaced aluminum disks hung in a bath of liquid helium from a torsion fiber. The period of oscillation of the pile of disks depended on the elastic constant of the torsion fiber and on the effective moment of inertia of the system of disks plus attached fluid. The spacings between the disks were small enough so that a fluid with the viscosity of liquid helium above the λ point would be completely carried along with the disks (at the periods of oscillation employed), and the disks were light enough so that the moment of inertia of this dragged along fluid was a measurable fraction of the total moment of inertia. There were of course edge effects arising from the continuous transition between the rotation of the fluid in between the disks and the resting fluid outside. Andronikashvili showed that these edge effects were only small corrections, and was therefore able to infer the moment of inertia of the fluid between the disks from his measurements. In accordance with Landau's prediction, this moment of inertia I turned out to be smaller than the classical value I_0, the ratio between them depending on temperature; as the temperature was lowered, more and more of the helium became superfluid, and eventually the concentration of normal fluid became so low that the moment of inertia of the dragged along fluid was immeasurably small compared to the moment of inertia of the pile of disks. In the region where the ratio (2.1) could be measured, the results were in excellent agreement with the normal fluid concentration inferred from measurements of the velocity of second sound.[12]

Andronikashvili's experiment was designed from the start to be as close as possible to the idealized "rotating bucket" experiment of Landau; the latter consists in setting a bucket filled with liquid helium in steady rotation, in a slow, isothermal manner, so that thermal equilibrium is maintained throughout the process, and measuring the torque necessary to bring about the final state. The time integral of the torque is the angular momentum of the final state, and the ratio of the angular momentum to the final angular velocity is the equilibrium moment of inertia. There is little doubt that Andronikashvili's pile of disks is equivalent to a set of very flat cylindrical containers, and that the height of the cylindrical "bucket" is of no great consequence, except insofar as it determines the time required before a normal fluid gets dragged along completely. The question is purely whether complete thermal equilibrium was attained. If we were to interpret Andronikashvili's result in terms of the Landau (or any other equilibrium) theory of superfluidity, we would say that thermal equilibrium was indeed established, and the superfluid did not move with the set of disks because the collisions which establish thermal equilibrium and maintain it act in such a way as to prevent the superfluid component from acquiring a vortex motion, i.e.,

$$\mathrm{curl}\mathbf{v}_s=0 \tag{3.1}$$

in thermal equilibrium.

However, we are now in a position to prove that this interpretation is incorrect. The radius of the disks used by Andronikashvili was 1.72 cm, and we know from Sec. 2 that the correlation length Λ of liquid helium is less than 10^{-2} cm. The moment of inertia I of the liquid in thermal equilibrium is given by Eq. (2.2). Hence, if thermal equilibrium had been established in Andronikashvili's experiment, his measured moment of inertia would have had to be within one percent of the classical value I_0. The fact that Andronikashvili found an apparent moment of inertia I appreciably less than I_0 thus proves that thermal equilibrium was *not* attained in his experiment, in spite of the fact that no frictional effects could be seen at the speeds used by him. The division of liquid helium into normal and superfluid according to (2.3) is therefore *not* possible in thermal equilibrium, but refers to a nonequilibrium state.

4. DISCUSSION OF OTHER THEORIES

Having proved that the superfluid state cannot be thermodynamically stable, let us now examine the arguments in the literature designed to prove the opposite. The London-Tisza school of thought give few detailed arguments, if any; equilibrium is assumed because no frictional effects have been observed between superfluid and normal fluid, at least at small relative velocities v_n-v_s. Another argument is based on a supposed analogy between superfluidity and superconductivity; since the superconducting state is known to be an equilibrium state, and since the two phenomena are in some sense analogous, it is concluded that the superfluid state must also be an equilibrium state. While analogies between superfluidity and superconductivity do exist,[17] the relationship is by no means that simple. Zilsel[2] attempts to prove equilibrium superfluidity on the Bose-Einstein gas model. The extent to which an ideal Bose-Einstein gas may be

[16] E. L. Andronikashvili, J. Phys. (U.S.S.R.) **10**, 201 (1946).

[17] M. R. Schafroth, this issue [Phys. Rev. **100**, 502 (1955)].

considered a superfluid has been discussed by us[5]; the discrepancy between our conclusions and Zilsel's stem from the fact that Zilsel attempts to impose a "macroscopic boundary condition" (fixed linear momentum per unit volume of the fluid) on a *part* of the fluid. This is inconsistent with the infinite correlation length of the ideal Bose-Einstein gas arising from the long wavelength of the condensed state. It is understandable, therefore, that Zilsel arrives at results inconsistent with elementary Galilean invariance; a discussion of linear motion of a container filled with fluid is possible[8] and shows that the fluid in thermal equilibrium always has the normal inertia with respect to such motions, independent of the statistics or any other detailed properties.

Landau's argument[3] for equilibrium superfluidity is more elaborate. He observes that creation of a single phonon at the wall of a container is forbidden by conservation laws unless the bulk velocity of the fluid exceeds the velocity of sound; creation of rotons at the wall is inhibited by the energy gap which Landau assumes to exist for vortex motions. Assuming these arguments to be correct, for the moment, what do they show? They indicate that the superflow is maintained by the absence of collisions, *not* that it is the thermodynamic equilibrium state established and maintained by the collisions. The conservation laws which prevent single phonon emission at a wall do not prevent double phonon emission, and this would lead to an eventual damping out of the superflow even on Landau's own theory. But as this is a second-order effect, Landau's arguments do lead us to expect a metastable superfluid state, and his subsequent calculation of the moment of inertia of the fluid in a rotating bucket is presumably applicable to this metastable state.

The separation of liquid helium into normal fluid and superfluid components must be understood as a division according to relaxation times. Certain fluid motions have relaxation times fast compared to the duration of usual experiments; these modes are combined into the "normal fluid"; other fluid motions, namely the bulk motions of semimacroscopic numbers of strongly correlated particles, have long relaxation times for exchange of momentum with the normal fluid or with walls; these motions are combined into the "superfluid." The argument between Landau[18] and Tisza[19] whether phonons should be considered part of the normal fluid or of the superfluid has since been decided experimentally in favor of including the phonons into the normal fluid.[12] We conclude that the phonon gas has fast relaxation times compared to the relaxation times of the modes of bulk motion of the fluid. This is consistent with theoretical estimates of mean free paths of phonons.[20] Tisza's argument was[9]: "only elementary excitations associated with definite masses can make it understandable that the liquid breaks up into two components." It is indeed hard to see how phonons, which can be created or destroyed with arbitrarily small amounts of energy, could be separated *in equilibrium* from the bulk motions of the fluid, But since the true separation is according to relaxation times, and does not persist in equilibrium, this argument is irrelevant.

[18] L. Landau, Phys. Rev. **75**, 884 (1949).
[19] L. Tisza, Phys. Rev. **75**, 885 (1949).
[20] L. Landau and I. M. Khalatnikov, J. Exptl. Theoret. Phys. (U.S.S.R.) **19**, 637, 709 (1947).

5. DISCUSSION OF AN EXPERIMENT

London[21] has suggested a crucial experiment to decide whether the superfluid state is thermodynamically stable, or merely metastable. This experiment consists in setting the bucket into rotation above the λ point, and then cooling the helium below the λ point while the bucket is left in free rotation. If the superfluid state is the thermodynamical equilibrium state, the superfluid component which is created during the cooling has no angular momentum, and the excess angular momentum is imparted to the bucket. Thus as one cools down, the angular velocity of the bucket must increase. If, on the other hand, the superfluid state is only metastable, no change in the angular velocity of the bucket is expected.

This experiment is very difficult to perform, especially since the required angular velocities are very small indeed. According to the ideas of H. and F. London, pure equilibrium superflow is expected only in that region of angular velocities in which we showed it to occur[5] for the ideal Bose-Einstein gas. These angular velocities are much too small for practical measurements. In view of our earlier arguments, there is no reason to expect equilibrium superfluidity even within these limits. But even if we allow appreciably larger angular velocities (e.g., similar to the ones used by Andronikashvili in his first experiment[16]) the experiment is still very difficult to perform.

In the experiments of Hollis-Hallett[22] and Osborne[23] with rotating buckets or sets of vanes frictional effects between superfluid and normal fluid were observed; however, the angular velocities involved were so large that these results are not necessarily in contradiction to equilibrium superfluidity at small values of $v_n - v_s$. These experiments have been discussed recently by Klemens.[6,10]

We shall devote our attention mainly to an experiment carried out by Andronikashvili.[24] A set of cylindrical vanes, similar to the one used in Andronikashvili's other experiment,[16] was suspended in liquid helium.

[21] H. London, Phys. Soc. Camb. Conference Report, 48, 1946 (Physical Society, London, 1947). The same experiment was suggested by us in our earlier Letter to the Editor, J. M. Blatt and S. T. Butler, Phys. Rev. **96**, 1149 (1954), in ignorance of the prior suggestion of London.
[22] A. C. Hollis-Hallett, Proc. Roy. Soc. (London) **A210**, 404 (1952).
[23] D. V. Osborne, Proc. Phys. Soc. (London) **A63**, 909 (1950).
[24] E. L. Andronikashvili, J. Exptl. Theoret. Phys. (U.S.S.R.) **22**, 62 (1952).

Initially both the helium and the vanes were at rest, and the temperature was slightly below the λ point. The system of vanes was then set into forced rotation and kept rotating at constant speed while the helium was cooled down to 1.5°K. At this temperature the vanes were stopped suddenly. The normal fluid concentration at this temperature is known (e.g., from the first Andronikashvili experiment) to be 12%. The system of vanes was then allowed to move freely, and the helium was warmed up quickly (in 15 seconds) to a temperature of 1.65°K, where the normal fluid concentration is 22 percent. Andronikashvili looked for a sudden deflection of the vanes, but failed to observe such an effect.

Assume first that the superfluid had been rotating along with the vanes during the cooling down period. When the vanes were stopped at 1.5°K, the normal fluid came to rest quickly. If the superfluid kept rotating for the next 15 seconds, while the system was warmed up to 1.65°K, then part of the (still rotating) superfluid would be transformed into normal fluid during the warming up, and this normal fluid would interact with the vanes, producing a deflection. Since no deflection was observed, one or more of the assumptions just made must have been wrong.

The simplest, but unacceptable, explanation of Andronikashvili's negative result is in terms of equilibrium superfluidity. If Eq. (3.1) holds, the superfluid is never set into motion at all. However, we know from general theoretical arguments that equilibrium superfluidity does not exist in liquid helium. We also know from other experiments[22,23] that the velocities used by Andronikashvili are so high that the superfluid is dragged along with the normal fluid quite quickly.

The most likely explanation of Andronikashvili's result is due to Klemens,[25] who points out that the relaxation time for exchange of momentum between superfluid and normal fluid, at the speeds used by Andronikashvili, is probably considerably less than the 15 seconds taken for warming up. Thus the superfluid did rotate along with the normal fluid during the cooling down period, but came very nearly to rest at the very beginning of the warming up period.

We shall now discuss what would happen if the same experiment were carried out at appreciably lower velocities. This turns out to be interesting because specific effects of the Bose-Einstein statistics come into play. Thus we feel that it would be very desirable to carry out the experiment, in spite of the great difficulties, in order to see to what extent simple arguments based on Bose-Einstein statistics are applicable to actual liquid helium.

In Bose-Einstein statistics of a weakly interacting gas, the probability of a transition *to* a single particle state i is proportional to N_i+1, where N_i is the number of particles already in state i. We shall discuss the limitation of working with single-particle states below; for the moment let us follow through the argument.

Suppose the experiment is started at a temperature *below* the λ point. Then there are already a macroscopic number of particles in the (nonrotating) ground state. Now we start rotating the vanes; another state ($m\neq 0$) becomes thermodynamically stable and if we wait long enough, the condensed particles transfer into that state. But by assumption the rate of rotation is slow enough so that thermal equilibrium is not reached, and the condensed particles stay in the $m=0$ state during the time of the experiment. However, what about the additional particles which drop from the normal fluid into the superfluid during the process of cooling down? In thermal equilibrium, these particles would condense into the most stable ($m\neq 0$) state in the rotating bucket. But the $m=0$ state has many more particles already in it than the true stable state. Hence, as other particles cascade down during the cooling, preferential condensation occurs into the $m=0$ state rather than into the true stable state. As more and more particles are attracted to the $m=0$ state, the preference for additional particles to go into it also increases. Hence we predict that the second Andronikashvili experiment, carried out under these conditions, would again show a negative result (no deflection of the vanes), this time because the superfluid would not be set into motion during the time of the experiment.[26]

On the other hand, no such argument applies if the experiment is started at a temperature *above* the λ point. Then the superfluid would be created for the first time during the cooling down of the already rotating bucket. The superfluid would be rotating all along, and (under conditions of low velocities and hence long relaxation times) the subsequent warming up would result in a deflection of the set of vanes.

If this analysis is correct, we have here one of the few experiments which could provide a clear experimental proof that the superfluidity of liquid helium (even at low velocities) is not an equilibrium phenomenon. For we have constructed two different states (superfluid rotating, and superfluid not rotating) at the same temperature and under the same external conditions (angular velocity of the vanes). Since the equilibrium state is unique, at least one of these states cannot be the equilibrium state, and it is then clear that long relaxation times play a dominant role in superfluid phenomena. This is contrary to the usual interpretation of the two-fluid theory, which assumes not only that the superfluid state is in thermodynamic equilibrium, but also that this thermodynamic equilibrium *is reached quickly* (compared to the times involved in experiments

[25] P. G. Klemens (private communication).

[26] In our earlier Letter to the Editor, J. M. Blatt and S. T. Butler, Phys. Rev. **96**, 1149 (1954), this argument was erroneously given as an explanation for Andronikashvili's actual experimental result, rather than as a prediction of a future experiment carried out at much lower velocities. We thank Dr. Klemens for calling our attention to the shortness of the relaxation time under Andronikashvili's actual experimental conditions.

with sets of vanes). It should be noted that this experiment does *not* involve a measurement of the actual relaxation time, and is hence feasible even if the relaxation time (at the velocities used) turns out to be very large indeed.

It remains to discuss the validity of working with single-particle states of the interacting Bose-Einstein gas. We have pointed out already[10] that a *straightforward* application of this model to actual liquid helium is not valid. However, there is reason to believe that a description of liquid helium in terms of equivalent "particles" (which obey Bose-Einstein statistics) is at least a good first approximation. The deviations from the simple-minded interacting Bose-Einstein gas model arise in two ways: the energy of these "particles" is no longer proportional to the square of the momentum, and the collision cross sections are modified considerably. The argument which we have given above has used neither the energy spectrum nor the values of collision cross sections, but merely the statistical factor N_i+1 for transitions into state i. Hence we believe that this particular application of the simple interacting Bose-Einstein gas model is valid, in spite of the fact that more detailed consideration based on this same model are open to grave doubt.

6. CONCLUSION

We have shown that the correlation length in actual liquid helium is shorter than the radius of the rotating "bucket" (set of vanes) used in the Andronikashvili experiment.[16] We could therefore apply the argument of a previous paper[8] on statistical mechanics of rotating buckets directly to liquid helium, the result being that no thermodynamically stable superfluid state can exist in that liquid. Since superfluid phenomena are in fact observed, the conclusion is that a superfluid nonequilibrium state with a long relaxation time is responsible for the observations.

A discussion of theoretical arguments given by others for an equilibrium superfluid state has shown that these arguments are invalid. The experimental situation is less satisfactory, in the sense that no unequivocal experimental test has been performed so far. The only available experiments have been carried out at velocities so high that frictional effects between superfluid and normal fluid could be expected even on an equilibrium theory of superfluidity. We have suggested a modification of a second experiment of Andronikashvili[34] which would allow a clear experimental decision between equilibrium and nonequilibrium superfluidity, as well as testing the validity of certain simple arguments based directly on the Bose-Einstein statistics.

No estimates of the relaxation time of the metastable superfluid state are contained in this paper, nor have we given any details about the nature of this state. We believe that it is too early to attempt a kinetic theory of liquid helium; the equilibrium properties of this substance should be understood first, and work along that line will be reported shortly.

ACKNOWLEDGMENTS

We would like to express our thanks to Dr. M. R. Schafroth who has contributed greatly to the contents of this paper through many valuable discussions. We are indebted to Dr. P. G. Klemens for valuable discussions on estimates of relaxation times, especially for disproving some incorrect arguments of ours at an early stage of this work. Last but not least, one of us (J.M.B.) would like to express his thanks to his old teacher and friend, Professor V. F. Weisskopf, for invaluable instruction in qualitative reasoning in theoretical physics, without which this work could never have been done.

Connection between Superfluidity and Superconductivity

M. R. SCHAFROTH
The F. B. S. Falkiner Nuclear Research and Adolph Basser Computing Laboratories, School of Physics, The University of Sydney, Sydney, N.S.W., Australia*

(Received April 4, 1955)

A general connection between equilibrium superfluidity and superconductivity is established. This connection allows us to apply to superconductivity a previously established theorem on the nonexistence of equilibrium superfluidity in actual physical systems. It follows that London's equations can only be approximately valid and have to be modified. An alternative equation proposed by Pippard is also in contradiction with the theorem.

1. INTRODUCTION

IT has long been suspected that there exist fundamental links between the two low-temperature phenomena of superfluidity and superconductivity.[1,2]

Recently, such a connection has become more apparent through the exhibition of a common model for the two phenomena: the perfect Bose-gas. It was shown that a gas of charged bosons below condensation is a superconductor.[3]

Furthermore, it has been shown that for a small range of angular velocities, the perfect Bose gas also exhibits superfluid properties.[4] It is the purpose of this paper to establish the validity of the connection between the two phenomena beyond the range of the Bose-gas model and to show that the two phenomena are related quite generally. More specifically, we shall define "perfect superconductivity" and "equilibrium superfluidity" of a physical system as being those properties which are actually exhibited by the Bose gas and then prove that for a system of identical spinless particles one of these properties entails the other. For systems of nonidentical particles cancellations can occur which destroy one effect, and therefore exceptions to the theorem are possible. Similar cancellations may occur in the case of particles with spin, namely when the contribution of the spins just cancels the contribution of the orbits. Such cancellations do not, however, affect the general connection between the two phenomena of superfluidity and superconductivity.

Beyond establishing the reason for the existence of a common model, the Bose gas, for the two phenomena, the above theorem yields important information about the superconductivity of real metals when viewed in connection with a recently established theorem on superfluidity in real physical systems.[5] It was pointed out there that for every real physical system, (i.e., a system of interacting particles) at a finite temperature there exists a "correlation length Λ" such that the momentum correlations of two particles farther apart than Λ decrease very rapidly. It was then proved[6] that no system with finite Λ, i.e., no real physical system, can show equilibrium superfluidity. In conjunction with the present theorem this entails that no superconducting metal can be a "perfect superconductor." It will be shown in Sec. 2 that a system obeying London's phenomenological equations[6] of superconductivity is a perfect superconductor. We can, therefore, conclude that the London equations are incompatible with the finiteness of the correlation length and can, therefore, only be approximately valid. Similarly, it can be shown that the modified equations proposed by Pippard[7] are inadmissible for the same reason. A more detailed discussion of this point will be given in a subsequent publication.

A word must be added at this stage about the question of how far one is justified in drawing from our theorem any conclusions about the behavior of actual superconductors. The possibility of accidental cancellations mentioned earlier presents no difficulty here, because the proof that the London equations are incompatible with the existence of a finite correlation length is independent of the actual occurrence of such cancellations. Apart, therefore, from the remote possibility that superconductivity be an effect of the spins rather than orbital motions—which seems to be excluded both by the existence of charge-carrying supercurrents and by the essentially diamagnetic nature of the Meissner-Ochsenfeld effect—our theorem can be applied to actual superconductors.

2. DEFINITIONS

We first define our terms precisely.

A. Consider a system in thermal equilibrium rotating with angular velocity ω around the z-axis and denote its angular momentum by $L(\omega)$. The quantity

$$I \equiv (dL/d\omega)_{\omega=0} \tag{1}$$

* Also supported by the Nuclear Research Foundation within the University of Sydney.

[1] J. G. Daunt and K. Mendelssohn, Nature **150**, 604 (1942).

[2] F. London, *Superfluids I* (John Wiley and Sons, Inc., New York, 1950), Introduction.

[3] M. R. Schafroth, Phys. Rev. **96**, 1149 (1954), and this issue [Phys. Rev. **100**, 463 (1955)].

[4] J. M. Blatt and S. T. Butler, this issue [Phys. Rev. **100**, 476 (1955)].

[5] Blatt, Butler, and Schafroth, this issue [Phys. Rev. **100**, 481 (1955)].

[6] See reference 2, paragraph 8.

[7] A. B. Pippard, Proc. Roy. Soc. (London) **A216**, 547 (1953).

is the moment of inertia. Its classical value is, for a system of N identical particles with mass M

$$I_0 = NM\langle x^2+y^2\rangle, \tag{2}$$

where $\langle\ \rangle$ denotes the average in thermal equilibrium.

Definition I: We call "equilibrium superfluid" a system for which $(I-I_0)/I_0$ does not decrease indefinitely as $\langle x^2+y^2\rangle$ is increased, i.e.,

$$(I-I_0)/I_0 \geq \eta > 0 \text{ for all } \langle x^2+y^2\rangle. \tag{3}$$

Note 1: It is known[4] that the ideal Bose gas is an equilibrium superfluid in this sense. It is also known[4] that equilibrium superfluidity is not enough to ensure actual superfluid behavior.

B. *Perfect Superconductivity*

Consider a system of charged particles in a fixed cylindrical volume V in a homogeneous magnetic field H parallel to the axis of the cylinder. Denote the total magnetic moment this system acquires in thermal equilibrium by $M(H)$. We shall define a quantity

$$\chi'(V) \equiv \frac{1}{V}\left(\frac{\partial M}{\partial H}\right)_{H=0}. \tag{4}$$

For any system with paramagnetic or diamagnetic behavior, χ' is independent of V.

Note 2: The quantity χ' here is not necessarily identical with the magnetic susceptibility χ as usually defined. The latter is

$$\chi = \lim_{H\to 0}\lim_{\Delta H\to 0}\lim_{V\to\infty}\left(\frac{1}{V}\frac{\Delta M}{\Delta H}\right), \tag{5}$$

whereas (4) means, for large volumes

$$\chi'(\infty) = \lim_{V\to\infty}\left(\frac{1}{V}\lim_{H\to 0}\lim_{\Delta H\to 0}\frac{\Delta M}{\Delta H}\right). \tag{6}$$

The limiting processes are not necessarily interchangeable, and so χ' and χ may turn out to be different. However, we are not concerned here with the value of the susceptibility; we only ask whether the last limiting process in (6) exists or not; i.e., whether χ' approaches a finite limit or whether it increases indefinitely with the dimensions of the container. This question, however, makes clearly no sense in relation to the susceptibility χ.

Definition II: A system for which

(a) $\chi' < 0$

(b) $-\chi'$

increases indefinitely, $\propto R^2$, with the radius R of the cylinder, shall be called a "perfect superconductor."

Definition IIa: A system for which

(a) $\chi' > 0$

(b) $\chi' \propto R^2$

shall be called a "perfect ferromagnet."

Theorem I: A system for which London's equation for the current density[5]

$$-\lambda c\ \mathrm{curl}\mathbf{i} = \mathbf{B} \tag{7}$$

holds, is a perfect superconductor. In actual superconductors, the only source of $\mathbf{B}$ is the supercurrent $\mathbf{i}$ itself, so that in Maxwell's equation

$$\mathrm{curl}\mathbf{H} = (4\pi/c)\mathbf{i} \tag{8}$$

one has to put $\mathbf{H}=\mathbf{B}$. (7) and (8) together, then, do not allow any homogeneous field to penetrate. However, as has been pointed out earlier,[3] it is a consistent "gedanken-experiment" to allow for a magnetic polarization $\mathbf{P}(x)$ inside the superconductor (arising e.g., out of a polarization of the nuclear moments) which has no other interaction with the superconducting particles than the magnetic one. The result is that, since now $\mathbf{B}=\mathbf{H}+4\pi\mathbf{P}$, a homogeneous $\mathbf{B}$ is possible, and (8) can be disregarded, its only function being to determine the necessary polarization $\mathbf{P}$.

For a cylindrical container with $\mathbf{B}$ parallel to the axis (7) at once yields

$$i_\theta = i_r = 0; \quad i_\varphi = -\frac{1}{\lambda c}\frac{r}{2}B. \tag{9}$$

Therefore,

$$\mathbf{M} = \frac{1}{2}\int d^3x(\mathbf{r}\times\mathbf{i}), \tag{10}$$

which yields

$$M_x = M_y = 0,$$

and

$$M_z = -(1/8\lambda c)BVR^2, \tag{11}$$

and therefore,

$$\chi' = -(1/8\lambda c)R^2. \tag{12}$$

This proves our assertion.

3. CONNECTION THEOREM

We wish now to prove the following theorem, which forms the main body of this paper:

Theorem II. If a system of identical spinless particles is an equilibrium superfluid, then the same system, when the particles are given a test charge ϵ, is also a perfect superconductor; and, conversely, any perfect superconductor consisting of identical spinless particles, is an equilibrium superfluid.

More in detail, we shall show that the moment of inertia I (1) and the quantity χ' (4) are related by

$$I = I_0\left(1+\frac{2\chi'}{n\langle x^2+y^2\rangle}\cdot\frac{Mc^2}{\epsilon^2}\right) \tag{13}$$

where n is the density of particles. (Note that χ' is proportional to ϵ^2, so that the test charge ϵ actually drops out.)

Note 3: The restriction to spinless particles may easily be removed, at least as long as one neglects any spin-orbit coupling. One can then define uniquely an orbital part of χ', χ_l' and a spin

[3] See reference 2, paragraph 3.

[5] M. R. Schafroth, Helv. Phys. Acta 24, 645 (1951).

part χ_1'. In terms of these one finds that (13) is replaced by

$$I = I_0\left[1+\frac{2(\chi_1'+g^{-1}\chi_2')}{n(x^2+y^2)}\frac{Mc^2}{e^2}\right], \tag{13'}$$

where g is the g-factor of the particles. This can lead to an exception from the theorem, when $\chi_1' \propto \langle x^2+y^2\rangle$, $\chi_2' \propto \langle x^2+y^2\rangle$, but $\chi_1'+g^{-1}\chi_2' = o(\langle x^2+y^2\rangle)$; i.e., when the system is at the same time a perfect superconductor and a perfect ferromagnet, in such a ratio that the two susceptibilities just cancel. This case is clearly too exceptional to warrant the complication of carrying through the proof of (13') in detail.

Note 4: Similarly, the restriction to identical particles is not very serious. Taking r kinds of particles with charges ϵ_i, masses M_i, densities n_i and g-factors g_i, (13') has to be replaced by

$$\frac{I-I_0}{I_0} = \sum_{i,j=1}^{r}\langle C_{ij}{}^{(1)}+C_{ij}{}^{(2)}\rangle,$$

$$\chi' = \frac{1}{2Vc^2}\sum_{i,j=1}^{r}\langle C_{ij}{}^{(1)}+g_ig_jC_{ij}{}^{(2)}\rangle\frac{\epsilon_i\epsilon_j}{M_iM_j}, \tag{13''}$$

where $C_{ij}{}^{(1)}$, $C_{ij}{}^{(2)}$ are certain coefficients. For the system to be either an equilibrium superfluid, a perfect superconductor or a perfect ferromagnet, at least one of the C's has to be $\propto\langle x^2+y^2\rangle$. It can, however, happen that for certain values of ϵ_i, m_i, g_i one of the two expressions $(I-I_0)/I_0$ and χ' is $\propto\langle x^2+y^2\rangle$, whereas in the other the corresponding terms just cancel. Accidental cancellations, therefore, are possible which invalidate the theorem.

We now proceed to prove (13). The Hamiltonian for N identical spinless particles of mass M and charge ϵ in a volume V under the influence of a homogeneous magnetic field H_z is

$$H = \sum_{i=1}^{N}\frac{1}{2M}\left(\mathbf{p}_i-\frac{\epsilon}{2c}\mathbf{H}\times\mathbf{r}_i\right)^2+V(\mathbf{r}_1,\cdots\mathbf{r}_N). \tag{14}$$

The magnetic moment operator is

$$\boldsymbol{\mu} = \sum_{i=1}^{N}\frac{\epsilon}{Mc}\mathbf{r}_i\times\left(\mathbf{p}_i-\frac{\epsilon}{2c}\mathbf{H}\times\mathbf{r}_i\right). \tag{15}$$

The average magnetic moment in thermal equilibrium is, therefore,

$$\mathbf{M} = \langle\boldsymbol{\mu}\rangle = \mathrm{Trace}\left\{\left[\sum_k\frac{\epsilon}{Mc}\mathbf{r}_k\times\left(\mathbf{p}_k-\frac{\epsilon}{2c}\mathbf{H}\times\mathbf{r}_k\right)\right]\right.$$

$$\left.\times\exp\left[-\beta\left(\sum_i\frac{1}{2M}\left(\mathbf{p}_i-\frac{\epsilon}{2c}\mathbf{H}\times\mathbf{r}_i\right)^2\right)+V\right]\right\}, \tag{16}$$

where the trace has to be extended over the space of all properly symmetrized states, and where $\beta = 1/kT$.

Introducing the Larmor angular velocity

$$\boldsymbol{\omega}_0 = e\mathbf{H}/2Mc \tag{17}$$

(16) can be written

$$\mathbf{M} = \frac{\epsilon}{MC}\mathrm{Trace}\left\{\left[\sum_k(\mathbf{r}_k\times\mathbf{p}_k-M\mathbf{r}_k\times(\boldsymbol{\omega}_0\times\mathbf{r}_k)\right]\right.$$

$$\left.\times\exp\left[-\beta\sum_i\left(\frac{1}{2M}(\mathbf{p}_i-M\boldsymbol{\omega}_0\times\mathbf{r}_i)^2+V\right)\right]\right\}. \tag{18}$$

On the other hand, the angular momentum in thermal equilibrium of the same system without magnetic field, but rotating with angular velocity $\boldsymbol{\omega}_0$, is given by taking the average of the angular momentum operator

$$\mathbf{L} = \sum_k\mathbf{r}_k\times\mathbf{p}_k \tag{19}$$

over a canonical ensemble with Hamiltonian[10]

$$H' = \sum_{i=1}^{N}\frac{1}{2M}\mathbf{p}_i^2+V(\mathbf{r}_1,\cdots\mathbf{r}_N)-\boldsymbol{\omega}_0\cdot\mathbf{L}, \tag{20}$$

or

$$H' = \sum_{i=1}^{N}\left[\frac{1}{2M}(\mathbf{p}_i-M\boldsymbol{\omega}_0\times\mathbf{r}_i)^2-\frac{M}{2}(\boldsymbol{\omega}_0\times\mathbf{r}_i)^2\right]$$

$$+V(\mathbf{r}_1,\cdots\mathbf{r}_N). \tag{20'}$$

One finds, therefore

$$\mathbf{L}(\omega) = \mathrm{Trace}\{\sum_k(\mathbf{r}_k\times\mathbf{p}_k)\exp(-\beta H')\}. \tag{21}$$

Comparing (21) with (18) we see that, neglecting terms of second order in ω_0, one has

$$\mathbf{M} = \frac{\epsilon}{MC}\mathbf{L}(\boldsymbol{\omega}_0)-\frac{\epsilon}{MC}\mathrm{Trace}\left\{\left[\sum_k\mathbf{r}_k\times(\boldsymbol{\omega}_0\times\mathbf{r}_k)\right]\right.$$

$$\left.\times\exp\left[-\beta\sum_i\frac{1}{2M}\mathbf{p}_i^2+V\right]\right\}. \tag{22}$$

Taking $\boldsymbol{\omega}_0$ to lie in the z-direction this reads

$$\mathbf{M} = \frac{\epsilon}{MC}\mathbf{L}(\boldsymbol{\omega}_0)-\frac{\epsilon N}{C}\boldsymbol{\omega}_0\langle x^2+y^2\rangle, \tag{23}$$

where $\langle\ \rangle$ is the average over the equilibrium distribution at $\boldsymbol{\omega}_0=0$. Differentiating with respect to $\boldsymbol{\omega}_0$ and using (1), (2), (4), and (17) this readily gives (13).

The essential step in this proof is the expansion of (16) and (21) in powers of $\boldsymbol{\omega}_0$, keeping only linear terms. This is only justified by our particular definition of I, (2) and χ', (4) which is restricted to finite volumes. The expansions in power of ω_0 will, as a rule, be justified only for very small angular velocities, $\omega_0<\omega_1\sim\hbar/MR^2$. Therefore this procedure cannot be used for computing the actually measured quantities χ, (5) and

$$I' = \frac{\Delta L(\omega)}{\Delta\omega}, \tag{24}$$

with

$$\Delta\omega\gg\omega_1.$$

However, for the purpose of this paper, the foregoing procedure is justified.

[10] See the detailed discussion of this in Blatt, Butler and Schafroth, this issue [Phys. Rev. **100**, 481 (1955)].

4. DISCUSSION

It has been shown[5] that no real physical system can be an equilibrium superfluid. The reason for this is that for all systems of interacting particles there exists a correlation length Λ which is an intrinsic property of the system, independent of the size of the volume, such that particles which are apart by a distance $>\Lambda$ have strongly decreasing momentum correlations.

The theorem of Sec. 3 together with the above statement that no equilibrium superfluid exists yields at once:

Theorem III: No perfect superconductors exist in nature.

In conjunction with the result of Sec. 2 that a system obeying London's equation is a perfect superconductor, this implies:

Corollary: No physical system can obey London's equation (7) exactly.

The ideal Bose-Einstein gas is not a physical system, since it does have an infinite correlation length[4]; indeed, it obeys London's equation.[3]

Note 5: The theorem of reference 5 is restricted to particles without spin and does, therefore, not exclude an anomalous moment of inertia I which is larger than the classical value I_0; this means that we cannot make a statement about ferromagnets similar to the one about superconductors.

Theorem II therefore leads us to the conclusion that London's equations can only be approximately valid for actual superconductors and that they have to be modified to comply with the finiteness of the correlation length Λ. We shall leave the task of finding an appropriate form of this modification to a later paper.

At present, we only wish to recall that the necessity for modifying London's equations has already been inferred by Pippard[7] from experimental results on the penetration depth with impurity contents. Pippard proposes the following equation [reference 7, Eq. (7)] to replace London's equation (7):

$$\mathbf{i}(\mathbf{x}) = -\text{const} \int_V d^3x' \frac{e^{-r/\xi}}{r^4} \mathbf{r}[\mathbf{r}\cdot\mathbf{A}(\mathbf{x}')], \qquad (25)$$

where $\mathbf{r}=\mathbf{x}-\mathbf{x}'$, ξ is a characteristic length of the material, and $\mathbf{A}$ is the vector potential of the field $\mathbf{B}$, gauged to $\text{div}\mathbf{A}=0$ and $(\mathbf{A}\cdot\mathbf{n})=0$ at the surface. Equation (25) is, however, also inadmissible from our present point of view since it still describes a perfect superconductor. This can be seen qualitatively by noticing that Eq. (25) expresses the current density $\mathbf{i}$ as a certain average of the vector potential $\mathbf{A}$ over a region of space of the dimension of the characteristic length ξ. If one considers now a cylinder of radius R in a homogeneous field H_0, then the vector potential $\mathbf{A}=\frac{1}{2}\mathbf{i}\times\mathbf{H}_0$ increases linearly from the axis to the surface of the cylinder. If the radius of the cylinder is large compared to ξ the current density increases essentially linearly, too, and therefore the argument in Sec. II for the London equations goes through essentially unchanged, leading to a $\chi' \propto R^2$.

Indeed, Pippard's modification of the London equation goes in the opposite direction from the one required by our theorem. The London equation is a strongly nonlocal relation between field and current. The current density is determined only by knowledge of the field $\mathbf{B}$ over the whole volume of superconductor. The fact that it can be brought into the local-looking form

$$\mathbf{i} = -(1/\lambda c)\mathbf{A} \qquad (26)$$

[where $\mathbf{A}$ is the vector potential in the same gauge as in (25)] only shows that the vector potential in this gauge is a nonlocal description of the magnetic field. Pippard's procedure of smearing (26) over small volumes tends to make the relation between current and field even less local, whereas the finiteness of the correlation Λ requires that the relation in question should be more local than (26), namely such that the current density is determined by the field distribution within a volume of order Λ^3.

We therefore conclude that London's equation (7) as well as Pippard's equation (25) are incompatible with a finite correlation length Λ. In a later paper we shall propose modified phenomenological equations for superconductors which are consistent with a finite Λ, and it will be shown then that Pippard's experimental facts become naturally understandable on this new basis, the correlation length Λ taking the role of his parameter ξ.

ACKNOWLEDGEMENTS

This work has been greatly influenced by many stimulating discussions and constructive criticisms of Dr. J. M. Blatt and Dr. S. T. Butler, whose interest is gratefully acknowledged.

Resonance Irradiation of Mercury Vapor in Nitrogen

J. A. Berberet* and K. C. Clark
Department of Physics, University of Washington, Seattle, Washington
(Received October 13, 1954; revised manuscript received July 15, 1955)

A study has been made of phenomena related to the metastable 6^3P_0 mercury atom appearing in a cell which contains nitrogen and saturated mercury vapor at 22°C and is irradiated with 2537 A ($6^1S_0-6^3P_1$) resonance emission from a cool mercury arc. Various phenomena reported individually by many investigators are observed simultaneously and with improved techniques so that coherent correlations can be drawn and new conclusions developed. The relative 6^3P_0 population produced from 6^3P_1 atoms by inelastic collision is measured by self-absorption of 4047 A ($6^3P_0-7^3S_1$); its dependence upon 2537 A intensity is linear and its dependence upon nitrogen pressure is given by theory. The 6^3P_1 population is quenched by nitrogen. With conditions of high gas purity it is possible to produce large currents of heavy ions, the probable mechanism being a three-body collision between 6^3P_0, 6^3P_1, and N_2 to form a highly excited molecule, which becomes ionized in a subsequent step. The ion formation is a volume process, varying with the second power of the 2537 A intensity. The continuum intensity near 4850 A is linear with 2537 A intensity and shows a pressure dependence consistent with molecular formation in a three-body collision between 6^3P_0, 6^1S_0, and N_2. Mean lifetimes of metastable population, ion formation, and continuum intensity are measured as a function of pressure of nitrogen. They are found to differ distinctly, being generally larger in the order named. At high irradiation intensity an ion current of 405 μa was observed, leading to a prediction that several percent of the mercury atoms could be in the metastable state.

INTRODUCTION

MERCURY atoms in the 6^3P_0 metastable state and phenomena associated with them have been the subject of many investigations.[1-3] Both electronic and resonance excitation have been used to produce high concentrations of these excited atoms. In the latter method, atoms raised to the 6^3P_1 state by 2537 A resonance radiation are transferred to the 6^3P_0 state in collisions of the second kind (see Fig. 1). Certain gases, particularly nitrogen, are very efficient in effecting this transition, the process having been studied early by Klumb and Pringsheim[4] and recently in detail by Matland.[5] Theories have predicted that approximately one out of every few hundred mercury atoms can be in the metastable state.[6-8] Foote[6] has predicted that as high as one out of every thirty-six could be in this state. Until the present work, high concentrations have not been experimentally measured. Buehl[9] did succeed in measuring metastable concentrations, but he used electronic excitation from which only very small concentrations were attainable.

The long mean life of mercury atoms in the 6^3P_0 state (of the order of milliseconds) has been measured by many workers.[6,7,10-16] Development of a rigorous theory predicting the mean life as a function of the pressure of nitrogen is complicated by diffusion of resonance radiation, transitions to the 6^3P_1 state and to the ground state, collision broadening, and wall effects. Nevertheless, semiquantitative theories have been able to predict the manner in which both the population of metastables and the mean life of this state vary as a function of the pressure of gas which is responsible for producing large populations of metastables.[12-14]

The population of the 6^3P_1 state in the presence of such a gas has likewise been frequently investigated.[4,17-24] It has been definitely established that the number of 6^3P_1 atoms is decreased with increasing nitrogen pressure. Collisions of the second kind producing 6^3P_0 atoms are chiefly responsible for this decrease. By the inverse of this transition, in decay the 6^3P_1 population decreases at the same rate as the 6^3P_0 population.[12,13,15]

The several continuous spectra in mercury discharges arise from molecular forms of mercury, although many of the assignments of molecular states are still somewhat in doubt. Certain of the continua have been observed in mercury vapor excited by resonance radiation, those near 4850 A, 3350 A, and 2537 A being most prominent. Their intensities are proportional to the intensity of the 2537 A source[25,26]; all three have been variously reported to have long decay times.[27] The continuum near 4850 A is probably basically due to the molecular combination of a 6^3P_0 atom with a 6^1S_0 atom, although evidence to the contrary has been offered.[25,28]

* Now at The General Electric Company, Hanford Atomic Products Operation, Richland, Washington.

[1] E. J. B. Willey, *Collisions of the Second Kind* (Edward Arnold and Company, London, 1937).
[2] M. W. Zemansky and A. Mitchell, *Resonance Radiation and Excited Atoms* (Cambridge University Press, London, 1934).
[3] J. A. Berberet, doctoral thesis, University of Washington, 1952 (unpublished).
[4] H. Klumb and P. Pringsheim, Z. Physik **52**, 610 (1928).
[5] C. G. Matland, Phys. Rev. **92**, 637 (1953).
[6] P. D. Foote, Phys. Rev. **30**, 288 (1927).
[7] E. Gaviola, Phil. Mag. **6**, 1154 (1928); **6**, 1167 (1928).
[8] R. W. Wood and E. Gaviola, Phil. Mag. **6**, 1191 (1928).
[9] A. Buehl, Helv. Phys. Acta **6**, 231 (1933).
[10] Asada, Ladenburg, and Tietze, Physik Z. **29**, 549 (1928).
[11] T. Asada, Physik Z. **29**, 708 (1928).
[12] M. L. Pool, Phys. Rev. **33**, 22 (1929).
[13] M. L. Pool, Phys. Rev. **38**, 955 (1931).
[14] E. W. Samson, Phys. Rev. **40**, 940 (1932).
[15] M. W. Zemansky, Phys. Rev. **29**, 513 (1927).
[16] M. W. Zemansky, Phys. Rev. **34**, 213 (1929).
[17] G. Cario and J. Franck, Z. Physik **37**, 619 (1926).
[18] E. Gaviola, Phys. Rev. **33**, 309 (1929).
[19] H. Kallman and F. London, Z. Physik. Chem. Abstracts **B2**, 207 (1929).
[20] R. Ladenburg, Z. Physik **48**, 15 (1928).
[21] J. S. Owens and O. S. Duffendack, Phys. Rev. **40**, 1042 (1932).
[22] H. A. Stuart, Z. Physik **32**, 262 (1925).
[23] M. W. Zemansky, Phys. Rev. **31**, 812 (1928).
[24] M. W. Zemansky, Phys. Rev. **36**, 919 (1930).
[25] Foote, Ruark, and Chenault, Phys. Rev. **37**, 1685 (1931).
[26] J. Franck and W. Grotrian, Z. Physik **4**, 89 (1921).
[27] Lord Rayleigh, Proc. Roy. Soc. (London) **A114**, 620 (1927).
[28] F. G. Houtermans, Z. Physik **41**, 140 (1927).

McCoubrey[29] has offered a detailed description of the emission process from this excited molecule in collision with two 6^1S_0 atoms. The presence of nitrogen makes possible the observance of the 4850 A band with resonance excitation at room temperature.[25,30]

Steubing[31] determined that ions were present in mercury vapor excited by 2537 A radiation. The most likely source of ionization energy is that available from a 6^3P_0 atom and a 6^3P_1 atom, which on combining lead to an ionized molecule. Using mass spectrograph techniques, Arnot and Millegan[32] have shown the ionization potential of the Hg_2 molecule to be 9.58 ev. Since the combined energy of the 6^3P_0 (4.67 ev) and 6^3P_1 (4.89 ev) atoms plus the heat of formation is 9.62 ev, the above mechanism is energetically possible. A combination of two 6^3P_0 atoms would fail to provide the ionization energy for the molecule, although subsequent thermal ionization of this molecule by an electron of about 0.2 ev is a possibility. Not even two 6^3P_1 atoms could furnish the 10.38 ev needed to ionize an atom, and the time constants of decay here reported discount the occurrence of ionization in a single collision process. Recent work by Biondi[33,34] with rare gases shows that collisions of two metastable atoms can produce atomic ions. In the case of resonance-excited mercury, however, the populated metastable levels lie well below half the the atomic ionization limit, and the molecular ion is therefore favored.

Other mechanisms of ion formation not related to the 6^3P_0 atom have also been postulated.[32,35] In a recent study, Biberman[36] attempts to show that this ion is not formed in the volume of the gas, but is due to electrons ejected from the electrodes by 6^3P_0 atoms. While it has been definitely established that 6^3P_0 atoms do eject electrons in collisions at the surface of a conductor,[37,38] numerous investigations, including the present one, have shown that the ions are formed in the volume of the vapor.[39–43]

Most of the phenomena relating to the metastable 6^3P_0 atom have in the past been studied individually. Consequently, much of the data presented could not be completely interpreted, or interpretations were based upon assumptions which were not substantially supported by experiments of others. It is the purpose of this work to present simultaneous correlations of many of these phenomena, leading to new information and a better understanding of the processes involved.

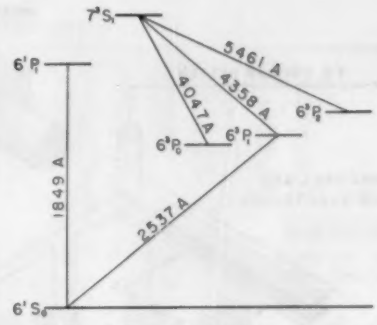


FIG. 1. Energy levels of mercury.

EXPERIMENTAL PROCEDURE

Resonance Cell

A resonance cell was designed to study the population of 6^3P_0 atoms by absorption of 4047 A radiation, to observe emission spectra from mercury vapor excited by 2537 A radiation, and to measure ion formation in the volume of the vapor. A schematic diagram of this cell along with auxiliary equipment is shown in Fig. 2. Commercial germicidal glass tubing of 2.5 cm diameter, which transmits approximately 65% of the 2537 A radiation and a negligible amount of 1850 A, was used for the walls of the cell; flat Pyrex windows were sealed at each end. Two plane-parallel platinum electrodes, 23 cm long and 2.2 cm wide, were placed lengthwide in the germicidal glass tube with separation of 0.465 ± 0.025 cm. Platinum, having a work function 0.8 volt greater that the excitation energy of the 6^3P_0 atom, was used as electrode material to minimize the ejection of electrons by photons and by metastables. However, at low pressures of N_2 a small current of ejected electrons was observed, indicating a lowering of the work function by impurities, presumably mercury.

In use this cell was shielded optically (and electrically) to allow 2537 A radiation to pass laterally through the region between the electrodes. For absorption studies, radiation from the 4047 A source passed between the electrodes parallel to the axis of the cell.

Gas Generator

Several investigators have shown the need for nitrogen of high purity when employing this gas to produce large quantities of metastable atoms.[44–46] Spectroscopic

[29] A. O. McCoubrey, Phys. Rev. **93**, 1249 (1954).
[30] R. H. Brown, doctoral thesis, University of Washington, 1950 (unpublished).
[31] W. Steubing, Physik Z. **10**, 787 (1909).
[32] F. L. Arnot and J. C. Millegan, Proc. Roy. Soc. (London) **A153**, 359 (1935).
[33] M. A. Biondi, Phys. Rev. **88**, 660 (1952).
[34] M. A. Biondi, Phys. Rev. **90**, 730 (1953).
[35] F. L. Arnot and M. B. McEwen, Proc. Roy. Soc. (London) **A171**, 106 (1939).
[36] L. M. Biberman and I. M. Gurevich, J. Exptl. Theoret. Phys. (U.S.S.R.) **20**, 108 (1950).
[37] A. Cobas and W. E. Lamb, Phys. Rev. **68**, 327 (1944).
[38] S. Sonkin, Phys. Rev. **43**, 788 (1933).
[39] M. A. Biondi, Phys. Rev. **82**, 453 (1951).
[40] M. A. Biondi, Phys. Rev. **82**, 566 (1951).
[41] M. A. Biondi, Phys. Rev. **83**, 653 (1951).
[42] G. F. Rouse and G. W. Giddings, Proc. Natl. Acad. Sci. **11**, 514 (1925).
[43] G. F. Rouse and G. W. Giddings, Proc. Natl. Acad. Sci. **12**, 447 (1926).
[44] K. F. Bonhoeffer and P. Harteck, *Grundlagen der Photochemie* (T. Steinkopff, Dresden and Leipzig, 1933).
[45] P. D. Foote, Phys. Rev. **29**, 609 (1927).
[46] G. A. Rosselot, Phys. Rev. **49**, 871 (1936).

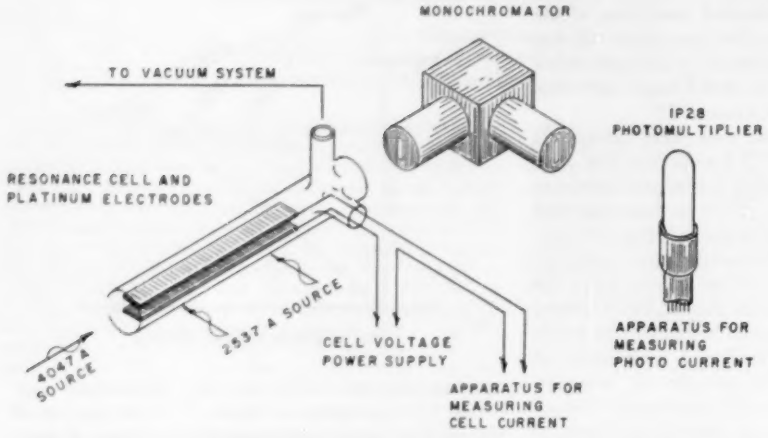


FIG. 2. Resonance cell for absorption and ionization studies.

traces of hydrogen and oxygen severely quench atoms in the 6^3P_0 state. Although water molecules have been shown to enhance the 6^3P_0 population, they may also be dissociated by 2537 A quanta, with the result that as H_2 and O_2 are accumulated, 6^3P_0 atoms are strongly quenched. Thus the N_2 to be used had to be free of H_2, O_2, and H_2O at least.

Nitrogen was prepared by the explosion of sodium azide in the generator shown in Fig. 3. The chief impurity anticipated in this process was H_2O vapor appearing originally as water of crystallization in the NaN_3 crystals. Interaction of this H_2O with free Na would give rise to free H_2. The generator system employed hot copper and copper oxide at 450°C to remove O_2 and H_2 respectively. Phosphorus pentoxide served to remove H_2O vapor, and the liquid air trap removed any volatile vapors as well as H_2O.

Prior to the explosion of the NaN_3, three bottles for collecting N_2 were pumped at a pressure less than 5×10^{-6} mm and baked at 450°C for 8 hours, after which time the open branch of each bottle was sealed off. The Cu and CuO at 525°C and the NaN_3 at 200°C were baked for 6 hours. The NaN_3 was then exploded with a cool torch. Nitrogen was then circulated for 48 hours over the P_2O_5 and hot Cu and CuO, and through the liquid air trap. Break-off tips in the other branch of the N_2 bottles were broken and the bottles sealed off after they had filled with N_2. With this preparation complete a bottle of N_2 could be sealed on the system designed for studying metastable atoms and N_2 admitted at the desired time by fracturing the remaining break-off tip.

The exceptionally large ion currents which were produced using this nitrogen and the constancy of this current when irradiated for extended periods with 2537 A radiation testify to its high purity. This high purity was achieved only after extended trials.

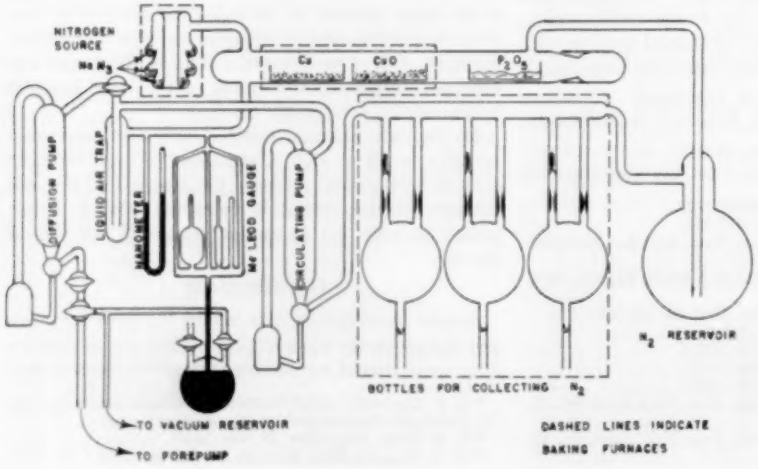


FIG. 3. Nitrogen generator.

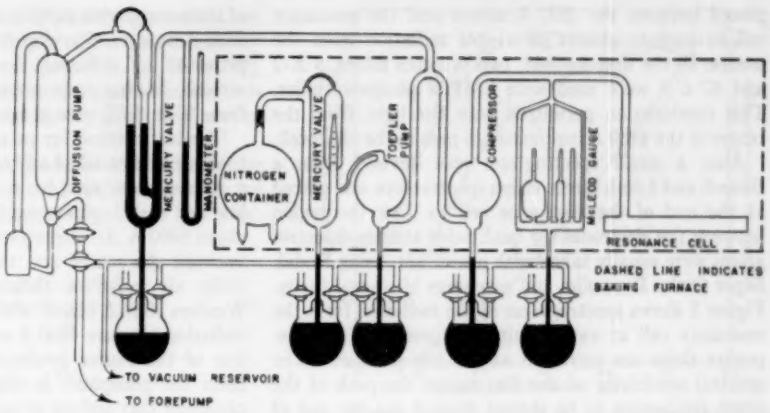


FIG. 4. Plan of apparatus.

Vacuum System

The vacuum system for the resonance cell, shown schematically in Fig. 4, used a mercury diffusion pump and a mechanical backing pump. Stopcocks were excluded from the system on the high vacuum side. Two mercury valves controlled by a vacuum reservoir were used to isolate different portions of the system. The mercury vapor concentration was set by saturation at 22°C room temperature throughout all experiments. In these studies the system was designed to permit a given sample of N_2 to be used for studying all the different phenomena. A Toepler pump was used to transfer N_2 from the storage container to the resonance cell and back. To attain highest pressures, a mercury "compressor" just preceding the resonance cell was used.

In assembling the system and purifying the mercury, great care was taken to prevent contamination. After a bottle of nitrogen was attached to the system, that portion enclosed in the dotted line in Fig. 4 was baked at 300°C for 12 hours with the pumps operating. At the conclusion of the baking, the valve to the diffusion pump was closed, and the system was observed to maintain a McLeod gauge pressure of less than 5×10^{-6} mm for 24 hours.

Light Sources

Light sources used for resonance excitation must operate at low temperatures to have narrow emission line widths and low self-reversal. The first source of 2537 A consisted of a pair of 15-watt commercial germicidal lamps, one on each side of the resonance cell. These were operated on direct current with resistive ballasts, and polarities could be reversed to prevent the "pumping" of mercury to the negative end. In a second type of source the same lamps were excited by 35-megacycle Hartley oscillators. A keying circuit could provide square wave pulses of light at the rate of 30 cps. The decay time of these square wave pulses was measured to be 0.2 millisecond.

For a source of 4047 A a Pyrex tube of $\frac{3}{4}$-in. diameter and 18-in. length, having a plane Pyrex window sealed on one end, was baked and filled with a drop of mercury and a few mm of argon. When excited by one of the rf oscillators and viewed end on, it was a small but intense source of 4047 A radiation. All sources were electrically shielded and air cooled and could be made to operate at constant intensity ($\pm1\%$) for periods of a few hours.

Measurement of 4850 A Continuum

The 4850 A continuum was studied by two methods. In both of these a Corning 9863 ultraviolet filter was

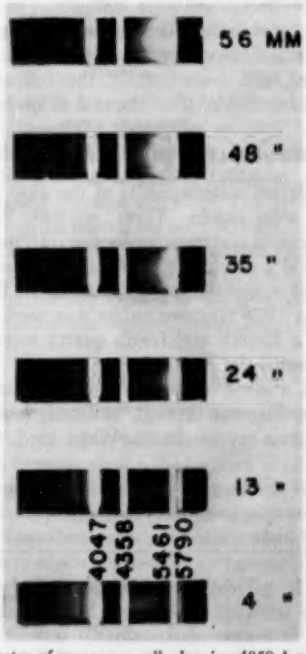


FIG. 5. Spectra of resonance cell, showing 4850 A continuum. Nitrogen pressures are labeled.

placed between the 2537 A source and the resonance cell to exclude almost all visible radiation from the source. In the first method, two Wratten filters, 8 *K*-2 and 47 *C*-5 were used with a 1P28 photomultiplier. This combination permitted only the light from the center of the 4850 A continuum to pass to the photocell.

Also, a small spectrograph was devised from a Bausch and Lomb direct vision spectroscope and placed at the end of the resonance cell to view the region between the electrodes for qualitative studies. Spectrograms were usually taken with a wide slit, using Kodak Super Ortho Press film and exposures of several hours. Figure 5 shows spectrograms of the radiation from the resonance cell at various nitrogen pressures. The exposure times are increased at the low pressures. The spectral sensitivity of the film causes the peak of the green continuum to be shifted toward the red end of the spectrum. The resonance cell emission lines are discussed in a later section.

Measurement of Resonance Cell Current

Steady-state resonance cell currents were measured with either a microammeter or a galvanometer, and periodically varying currents were measured by oscilloscope. A variable power supply was used to apply a voltage across the platinum electrodes. For a given nitrogen pressure and constant 2537 A radiation there was a pleateau region of collection voltage, where all ions being formed were drawn to the electrodes without appreciable recombination, diffusion, or additional ionization by collision. The current at this saturation voltage is thus proportional to the rate at which ions are formed, and consequently the cell current will frequently be referred to as the rate of ion formation.

Measurement of 2537 A Radiation

Only relative measurements of the 2537 A radiation intensity were made. Three methods were used: monochromator and photomultiplier; calibrated neutral screens; and a combination of filters, ultraviolet sensitive phosphor, and photomultiplier. In the first of these methods, a 1P28 photomultiplier was used in conjunction with a Bausch and Lomb quartz monochromator set to transmit the 2537 A line.

Neutral screens were made, using parallel black nylon monofilament threads, uniformly separated and mounted on a special frame. When used, the threads of the screens were parallel to the resonance cell and the 2537 A sources. Varying degrees of transmission were provided by different thread size and spacing. Six screens, having transmission values varying in steps from 18 to 88% were constructed and calibrated.

The last method for measuring the 2537 A intensity used a commercial phosphor (General Electric) having a narrow absorption band in the region of the 2537 A line and emitting between 5200 A and 5750 A, peaking about 5400 A. An ultraviolet filter, Corning 9863, placed between the source and the phosphor, cut out essentially all radiation above 4200 A. A second filter, Wratten 58 *B*-2, placed after the phosphor, transmitted radiation between 4800 A and 6400 A. Only the 2537 A line of the source produced any significant emission from the phosphor. It was only the light from the phosphor that arrived at the photomultiplier. Since the intensity of the emitted light was found to be proportional to the intensity of the 2537 A line, the photomultiplier current was proportional to the latter. This system was used mainly to monitor the 2537 A sources to insure constant intensity.

TABLE I. Absorption constant at various pressures.

Pressure of N_2 in mm	Absorption constant A_6
6.7	0.38
13.6	0.38
20.5	0.37
28.6	0.37
37.4	0.38
55.	0.38

Measurement of 4047 A Absorption

As shown in Fig. 2, radiation from the 4047 A source passed through the resonance cell to the monochromator, set to pass only the full 4047 A line, and then to the photomultiplier. The percent absorption was obtained by dividing the difference between 4047 A intensities with the 2537 A source first off and then on by the 4047 A intensity with the 2537 A source off. When the 2537 A source was operated in a square wave fashion, the absorption of 4047 A radiation varied as a function of time due to the time decay of the 6^3P_0 atoms. In this instance, the oscilloscope pattern of output of the photomultiplier showed the manner in which the absorption varied from its maximum steady state value to zero. To obtain the percent absorption, a synchronous motor was used to drive an interruptor synchronized so as to interrupt the 4047 A beam briefly during the off time of the 2537 A radiation; and from the resulting pattern it was simple to calculate the percent absorption as before.

OBSERVATIONS

Population of 6^3P_0 Atoms

The percent absorption is proportional to the concentration of 6^3P_0 atoms when the concentration is small. To obtain a more widely useful relation between these two quantities, we postulate that the 4047 A radiation is absorbed according to Beer's law and that the 4047 A emission line of the source is broad compared to the absorption line of the metastable atoms of the resonance cell. For this moderate absorption one may approximately represent the intensity distribution in the transmitted line by a broad, flat-topped distribution with a narrow, partly absorbed area at the

center. If the concentration of absorbing atoms is not too large, the absorption line is of fixed form and the absorbed area is simply proportional to $(1-e^{-kN_0})$, where N_0 is the concentration of metastables and k is proportional to the central absorption cross section per atom.

If the observables R and $\bar{R}$ are the total areas of the transmitted line with and without the presence of absorbing metastables, respectively, then the fractional absorption A is

$$A=\frac{\bar{R}-R}{\bar{R}}=\frac{K(1-e^{-kN_0})}{\bar{R}} \tag{1}$$

or,

$$A=A_0(1-e^{-kN_0}),$$

leading to

$$N_0=\frac{1}{k}\ln\frac{A_0}{A_0-A}. \tag{2}$$

The validity of the above assumed conditions of moderate, narrow absorption can be tested by checking the constancy of A_0 as N_0 is varied by known fractions over the full range of operation. The relative values of N_0 were easily provided by calibrated screens on the 2537 A source and use of the fact that N_0 is proportional to the intensity of the 2537 A. Equation (2) could then be solved for A_0 using known values for N_0 and A. Not only were semilogarithmic plots of Eq. (2) linear, but the value of A_0 at different N_2 pressures remained constant, indicating negligible effects of pressure broadening. The values of A_0 determined for six different pressures of N_2 are shown in Table I. Their constancy justifies the use of Eq. (2) up to $A=24\%$, the highest absorption value used in calculations. The facts that A_0 is considerably less than unity and that A reached only 0.73 even with the very high N_0 present during the production of the 405 μa current, as discussed later, give support to the above assumption of an emission line broader than the absorption line.

This absorption technique was used to measure the relative concentration N_0 of metastables as a function of the nitrogen pressure P in mm at constant 2537 A intensity. These data are given in Fig. 6(A), which shows the experimental points for N_0 *vs* P. Klumb and Pringsheim[4] have discussed the factors affecting N_0 as a function of P. They obtained the expression

$$N_0=MP/(1+aP+bP^2), \tag{3}$$

where M is a convenient scale factor and a and b are constants having physical significance but arbitrary values. If one uses $M=7.8$ and the values $a=0.0143$ and $b=0.00093$, this equation fits the present data very well, as shown by the curve in Fig. 6(A).

A low intensity of 2537 A radiation was used in order to avoid high absorption of 4047 A. The data obtained here are similar to those obtained by other investigators.[4,15,47] Further agreement with the work of others was obtained when it was found impossible to detect any absorption of the 4358 A and 5461 A lines, indicating that the populations of the 6^3P_1 and 6^3P_2 states were small.

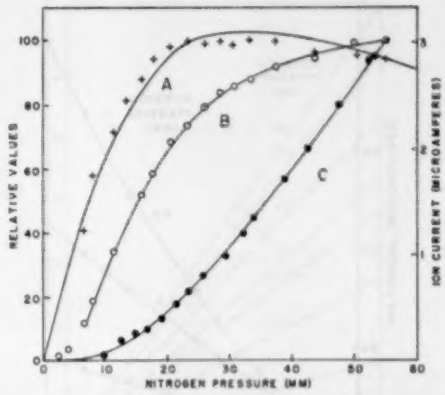


FIG. 6. Pressure dependence of resonance phenomena at constant 2537 A intensity. Curves show fit of theory to experimental points. A: Population of 6^3P_0 atoms. B: Rate of ion formation. C: Intensity of 4850 A continuum.

If the 4047 A absorption source was removed, it was possible to observe line emission excited by double absorption, as illustrated by the spectra of Fig. 5. The 6^3P_0 atoms, derived from resonance-excited 6^3P_1 atoms, were raised to the 7^3S_1 state by the 4047 A radiation which weakly penetrated the ultraviolet filter of the 2537 A source. The lines 5461 A, 4358 A, and (in part) 4047 A come from this excitation. The yellow line is presumably 5790 A from $(6^1P_1-6^1D_1)$, similarly excited by the $(6^3P_0-6^3D_1)$ absorption of 2967 A, which the ultraviolet filter readily transmits. The resonance cell line at 4047 A is due in part to radiation of this wavelength passing through the filter. The 5461 A line was shown not to be caused by such leakage, for the variation of its intensity with N_2 pressure was approximately the same as that of N_0. Because of the overlap of the 4850 A continuum, especially at high pressures, the 5461 A emission was not used for quantitative measurements of N_0.

Ion Formation

Ions were observed in the mercury vapor at all pressures of nitrogen with resonance excitation. That these ions were not due to the ejection of photoelectrons from the electrode surfaces was shown by the great reduction in current at low nitrogen pressures, by studies of transients, by the inability of a high intensity iron arc to produce measurable surface emission, and by supplementary experiments consisting of irradiating selected portions of the volume between widely separated electrodes. Characteristic curves of rate of ion

[47] E. Gaviola, Phys. Rev. **35**, 1226 (1930).

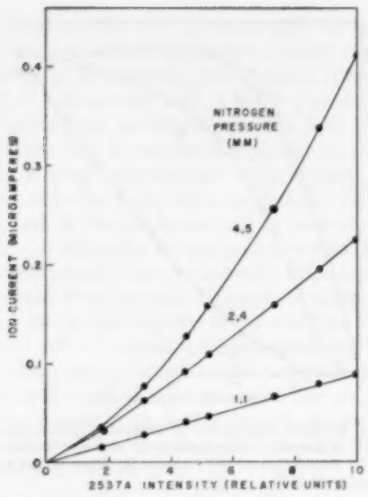


FIG. 7. Dependence of rate of ion formation on the first power of 2537 A intensity at low nitrogen pressures. Curves indicate experimental points.

formation *versus* cell voltage were taken at several nitrogen pressures between 1 and 55 mm. At all pressures plateaus were found to exist, and collection voltages in other measurements were always chosen on these plateaus to insure that negligible ionization by collision was occurring.

As shown in Figs. 7 and 8, the rate of ion formation exhibits a first power dependence upon the 2537 A intensity at a pressure of 1.1 mm. At pressures of 7 mm and higher the rate of ion formation is proportional to the second power of the intensity of the 2537 A radiation.

These results agree with those of other investigators.[41,43,48] At very low pressures metastable atoms, the number of which is proportional to the intensity of the 2537 A radiation, rapidly diffuse to the electrode surface and eject electrons. At higher pressures this process becomes unimportant compared with that involving collisions of two excited atoms within the volume. This process exhibits a square law dependence, and the most likely mechanism is

$$\begin{aligned} &Hg(6^3P_1)+N_2 \rightarrow Hg(6^3P_0)+N_2' \\ &Hg(6^3P_0)+Hg(6^3P_1)+N_2 \rightarrow Hg_2''+N_2 \qquad (4) \\ &Hg_2'' \rightarrow Hg_2^+ + e. \end{aligned}$$

Evidence in support of the postulated intermediate excited molecule is presented later in connection with studies of decay transients. The possibility suggested by Biondi[34] of single step ionization in a collision of two excited mercury atoms is not contradicted by these decay studies, as the necessary 6^3P_2 atoms would be much more abundant following his electrical excitation than after our resonance absorption.

[48] F. G. Houtermans, Z. Physik **41**, 619 (1927).

The rate of ion formation as a function of N_2 pressure, the 2537 A intensity being constant, is given in Fig. 6(B). The process involving the combination of a 6^3P_0 and a 6^3P_1 atom to form the ion is the only one which adequately explains these data. Since a three-body collision is required to satisfy the conservation of momentum and energy, the current I is proportional to N_0N_1P, where N_1 is the population of 6^3P_1 atoms. Since N_1 was not measured as a function of the pressure of nitrogen, it was assumed that the population varied in the manner described by the Stern-Volmer theory of quenching,[2]

$$N_1=[(N_1)_{P=0}]/(1+cP).$$

Thus

$$I=BN_0P/(1+cP). \qquad (5)$$

Using the values of N_0 given by a smooth curve through the experimental data of Fig. 6(A), and choosing the scale constant $B=0.0595$ and the quenching coefficient $c=0.0384$ mm^{-1}, one obtains a good fit of this equation to the experimental data, as demonstreated by the curve of Fig. 6(B). Close agreement at pressures below 7 mm is not expected, as it has just been shown that the postulated three-body collision process does not predominate in this region. It is seen that at 50 mm N_1 is reduced to 34% of its initial value. This amount of quenching is not inconsistent with other reports. Zemansky's theory[2] predicts a value of 25% at this pressure. Also, Badger[49] determined that the intensity of the scattered 2537 A radiation, while not directly proportional to the 6^3P_1 population, decreased at 50 mm of N_2 to about 30% of its value at zero pressure. The recent quenching cross section given by Matland[6] can be used in calculating an approximate quenching coefficient for the present experiment, following his method appropriate to a long, circular tube. One must use an equivalent radius for our electrode geometry. Using his cross section of 0.4×10^{-16} cm^2 and the extreme choices of equivalent radius as 1.3 cm and 0.23 cm, one finds that our observed value of $c=0.0384$ mm^{-1} is bracketed by the calculated values 0.013 mm^{-1} and 0.13 mm^{-1}, respectively.

The success of this mechanism in explaining the data lends added support to substantiate the hypothesis of Houtermans,[48] Snavely,[50] and Arnot and McEwen,[51] namely, that the ion arises from the combination of an atom in the 6^3P_0 state with one in the 6^3P_1 state in a three-body collision.[52]

[49] R. M. Badger, Z. Physik **55**, 56 (1929).
[50] B. L. Snavely, Phys. Rev. **52**, 174 (1937).
[51] F. L. Arnot and M. B. McEwen, Proc. Roy. Soc. (London) **A165**, 133 (1938).
[52] This theory in its present form fails to explain the reported dependence of ion formation on the third power of the 2537 A intensity at high vapor pressures.[6] However, it has been demonstrated in this work that high concentrations of 6^3P_0 atoms are possible. It is also possible that the cross section of interaction with an excited atom might be much greater than with an unexcited atom. The cubic law could thus conceivably arise at the higher vapor pressures if the 6^3P_0 atoms become so abundant that they commonly act as the third body.

It is well to note here that Houtermans observed a linear relationship between the rate of ion formation and the pressure of argon which was used as a diluent gas. It is fairly well established that argon does not appreciably quench 6^3P_1 atoms. Thus Eq. (5) would reduce to

$$I = BN_0P.$$

But Klumb and Pringsheim, using argon as a diluent gas, found that above 3 mm of argon, N_0 was a constant. Thus above 3 mm of A, I would be proportional to P only, and Houtermans' observations would be quite understandable.

A direct observation of the presence of 6^3P_0 atoms in the ion formation process was afforded by decreasing N_0 independently by absorption of 4047 A. One 2537 A source was operated continuously; the other was keyed by a 30-cps square wave, filtered to give only visible wavelengths, and used as the source of 4047 A radiation. The supply of 6^3P_1 atoms was thus held constant, and a small fraction of 6^3P_0 atoms was selectively removed by absorption. A resulting decrease in cell current, which was determined to be less than 0.2%, was clearly observed by oscilloscope. However, to use this method for distinguishing directly between ion processes involving either two 6^3P_0 atoms or one 6^3P_0 and one 6^3P_1, one would need to produce reductions in I of the order of 10%.

An ion current of 405 μa was observed at 50 mm pressure when the 2537 A intensity was temporarily made very large, and the absorption of 4047 A by metastables became 73%. In this experiment and those of other investigators it is reasonable to assume that at least one 6^3P_0 atom is required for every ion. The ratio of population of 6^3P_0 atoms to that of Hg atoms in the ground state N' is given by

$$N_0/N' \geqq I\tau_0/N'e, \tag{6}$$

where e is the electronic charge, τ_0 is the mean life of the 6^3P_0 state, and I is the cell current of 405 μa. For a temperature of 22°C N' is 5×10^{13} atoms/cm^3 and τ_0 at 55 mm of N_2 is measured (Fig. 13) to be 0.59 millisecond. Substituting these values gives

$$N_0/N' \geqq 1/700. \tag{7}$$

Thus at least 1 out of every 700 mercury atoms was in the 6^3P_0 state.

It is probable, however, that the estimate of one 6^3P_0 atom for every ion formed is conservatively low, for all studies concerning the mechanisms whereby 6^3P_0 atoms decay consider the number contributing to the formation of ions as being negligible. The exponential character of the decay of 6^3P_0 population illustrated in Fig. 10 shows that in decay the disappearance rate must predominantly have terms in the first power of N_0, not the N_0^2 or N_1N_0 terms characteristic of loss by ion formation. It is also apparent that in normal steady state measurements the process of ion formation is not dominant, for both the rate of production of 6^3P_0 atoms and their population are proportional to the 2537 A intensity. It is likely that loss by ion formation is still minor even at the increased intensity of irradiation used to produce the 405 μa current. Moreover, an intense 2537 A arc was not used here to produce this current. Thus it is quite conceivable that under favorable conditions as many as 1 out of every 20 mercury atoms could be in the 6^3P_0 state. The present evidence and its extended implication therefore give an approach to substantiating the theories of Wood and Gaviola[7,8] and of Foote,[6] which have predicted these high concentrations.

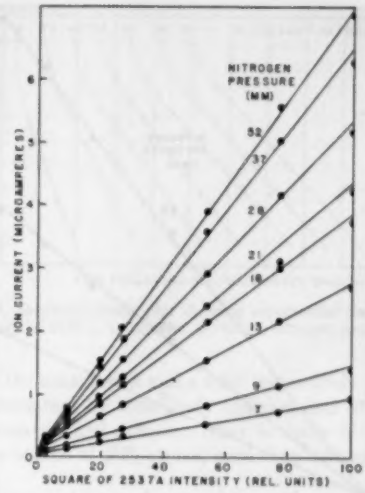


FIG. 8. Dependence of rate of ion formation on the second power of 2537 A intensity at high nitrogen pressures. Curves indicate experimental points.

Continuum at 4850 A

The central region of this 4850 A continuum was isolated by filters previously described and its intensity measured by photomultiplier. The direct proportionality between this intensity and that of the incident 2537 A radiation is shown in Fig. 9 at three nitrogen pressures. Since the population of metastables N_0 is proportional to the 2537 A intensity, these results agree with the following excitation process, leading to the emission of the 4850 A continuum:

$$\begin{aligned} &Hg(6^3P_1)+N_2 \rightarrow Hg(6^3P_0)+N_2' \\ &Hg(6^3P_0)+Hg(6^1S_0)+N_2 \rightarrow Hg_2'+N_2. \end{aligned} \tag{8}$$

Assuming that the population N_φ of these excited molecules is proportional in a steady state to the intensity of the 4850 A continuum and that the two main sinks for these molecules are radiation and diffusion to the walls, one equates rates of production and loss and

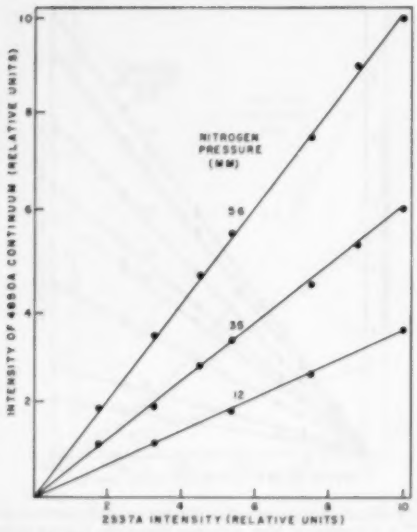


FIG. 9. Linear dependence of 4850 A continuum intensity on 2537 A intensity at three nitrogen pressures.

obtains

$$CN_0P=N_e(1+d/P)$$

or,

$$N_e=CN_0P^2/(P+d), \tag{9}$$

where C and d are constants and P is nitrogen pressure in mm. The dependence of N_0 on P is given independently in Fig. 6(A). The curve passing through the experimental points in Fig. 6(C) illustrates the success with which this equation can be fitted to the data, using $C=0.073$ and $d=155$. A small, constant intensity due to the measured background of glass fluorescence has been subtracted from the observed data.

McCoubrey[39] has shown evidence that in pure mercury vapor at 200°C the continuum is emitted as collision-induced radiation from the excited molecule on striking two ground-state atoms. No such collision process appears called for in the present studies, however, where nitrogen serves as collision particles.

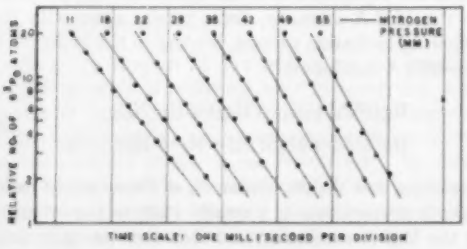


FIG. 10. Semilogarithmic plot showing exponential decay of population of 6^3P_0 atoms at various nitrogen pressures.

Effects of Large Voltage

Large changes in collection voltage caused noticeable effects, which will be reported for a pressure of 55 mm. The absorption of 4047 A decreased from 22.9 to 22.2% as the voltage changed from 0 to 400 volts, while the intensity of the 4850 A continuum increased linearly from 83 to 100%. The ion current, reaching the knee of the plateau region at about 40 volts, increased linearly from 90 to 100% between 40 and 400 volts.

Time-Dependent Transients

If the source of 2537 A is operated steadily and the collecting voltage is switched on and off as a square wave, one observes characteristic transients in the cell current. An oscillographic study of the ion pulse shape after the collecting voltage is turned on shows a linear decrease of ion density from the recombination-limited, free-field value to the steady value limited by the rate of production. To assume linearity of slope, it was necessary to avoid high ion densities, where appreciable recombination introduced curvature. In agreement with the processes involved, the latter steady value was proportional to the second power of the 2537 A intensity, and the former initial value was proportional to the first power. While the resonance cell was not designed with a uniform field E for accurate mobility measurements, the heavy ion mobility was measured to be 2.2 ± 0.2 cm²/volt sec over the observed range of E/P from 3.0 to 6.6 volts/cm mm. Mitchell and Ridler[38] have determined the mobility of Hg^+ in nitrogen to be 2.20, while for Hg_2^+ they give the theoretical value of 1.50 with the comment that it may be too high. Thus, the observed mobility in the resonance cell would somewhat favor the atomic ion. However, energy considerations and the work of many investigators require that the Hg_2^+ form be given preference over the Hg^+ form. The presence of positive ions in the volume of the resonance cell, as shown in these transients, discounts Biberman's contention that currents produced by resonance excitation are due only to ejection of electrons by metastables.

If the 2537 A source is keyed on and off as a square wave, additional information can be gained from observing the decay transients of the 6^3P_0 atom population, the cell current, and the 4850 A continuum. For the first named transients, the 4047 A source was operated steadily, and for the second, a constant collecting voltage was supplied. These decays are all essentially exponential, having lifetimes τ_0, τ_i, and τ_e, respectively. The switching time of the optical excitation and the collection time of the ions were both very rapid compared to the decay times observed. The results of the decay observations of these three quantities are graphed in Figs. 10, 11, 12, and 13. All decay curves commence at the end of the 2537 A radiation.

[38] J. H. Mitchell and K. Ridler, Proc. Roy. Soc. (London) **A146**, 911 (1934).

The lifetime τ_0 of the 6^3P_0 state as a function of pressure of nitrogen has been the subject of many investigations. It is well established that the decay is essentially exponential,[14] and we may write the time dependent concentration as

$$n_0 = N_0 e^{-\beta t}. \tag{10}$$

Samson has considered the diffusion of resonance radiation and the quenching of and radiation from 6^3P_1 and 6^3P_0 states in arriving at the following expression, which fits his experimental measurements:

$$\beta = \frac{1}{\tau_0} = \frac{f}{P} + \frac{gP + hP^2}{P + m}, \tag{11}$$

where f, g, h, and m are constants depending on the temperature and geometry of the resonance cell. The maximum lifetime of about 3 milliseconds observed by Samson in a cell whose dimensions exceeded 2 cm is reasonably matched by the value of about 0.9 millisecond in the present cell, whose smaller dimensions

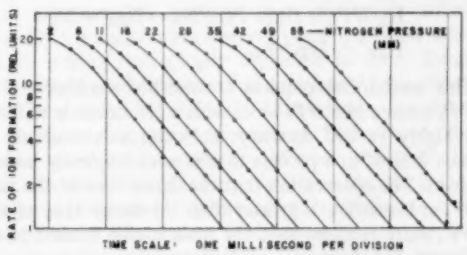


FIG. 11. Semilogarithmic plot showing exponential decay of rate of ion formation at various nitrogen pressures.

would enhance the loss by diffusion. The expected decrease of τ_0 with sufficiently large pressures is also observed.

One may refer to the observed cell current as the rate of ion formation because the transit time of the collected ions was always less than 5% of the current decay time. After about 0.5 millisecond the transients in Fig. 11 show a simple exponential decay, of lifetime τ_i, following a slower decay during the first few tenths of a millisecond. It is at once clear that since τ_i is greater than $\tau_0/2$ and in fact is slightly greater than τ_0, as shown in Fig. 13, then the ion cannot arise directly from the collision of two excited atoms and a third body. An intermediate excited molecule may be formed, whose lifetime is approximately τ_i and which then produces an ion by a subsequent process of unspecified nature, not involving excited atoms. The electronic state of this postulated molecule and the cause of its surprisingly long lifetime remain undetermined.

The decay of the concentration n of these highly excited molecules follows the equation

$$dn = \alpha n_0 n_1 dt - \mu n dt, \tag{12}$$

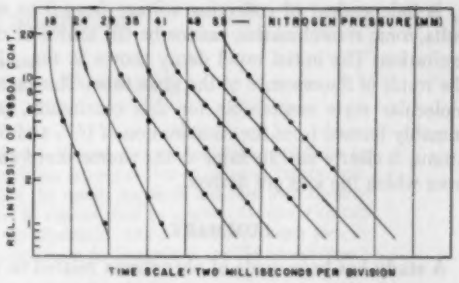


FIG. 12. Semilogarithmic plot showing exponential decay of intensity of 4850 A continuum at various nitrogen pressures.

where the constants α and μ refer respectively to molecule formation by three-body collision and to loss by ion formation and diffusion. Since in decay 6^3P_0 atoms are the source of 6^3P_1 atoms, each then decays as $e^{-\beta t}$, giving

$$dn = \alpha N_0 N_1 e^{-2\beta t} dt - \mu n dt. \tag{13}$$

The solution for this equation is

$$n = \frac{\alpha N_0 N_1}{(2\beta - \mu)} (2\beta e^{-\mu t} - \mu e^{-2\beta t}), \tag{14}$$

which has an initial slope of zero and which asymptotically approaches a simple exponential decay of lifetime $1/\mu$. The decay curves of Fig. 11 fit this analysis. As an independent check on Eq. (14), one may make a rough calculation of $\tau_0 = 1/\mu$ based on Eq. (14) and the small differences between these decay curves and their asymptotes. The values of τ_0 thus obtained range from 0.6 ± 0.2 to 0.5 ± 0.2 millisecond over the full pressure range and are in fair agreement with the τ_0 values in Fig. 13(A) as measured directly for the 6^3P_0 atoms.

The decay of the 4850 A continuum shows a lifetime τ_ϕ which increases greatly with pressure and reaches a value 3 to 5 times that of τ_0 and τ_i. See Fig. 13. Since

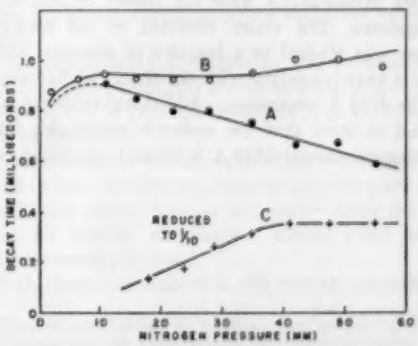


FIG. 13. Decay times of resonance cell phenomena at various nitrogen pressures. A: Population of 6^3P_0 atoms. B: Rate of ion formation. C: Intensity of 4850 A continuum.

τ_ϕ is independent of collecting voltage from 0 to 400 volts, ionic recombination cannot be the source of this excitation. The initial rapid decay shown in Fig. 12 is the result of fluorescence of the glass tube. The excited molecular state responsible for this continuum, presumably formed from the combination of 6^3P_0 and 6^1S_0 atoms, is clearly not the same as the shorter-lived state from which the ions are derived.

SUMMARY

A study has been made of phenomena related to the metastable 6^3P_0 mercury atom appearing in a resonance cell containing mercury vapor and nitrogen and irradiated with 2537 A from a cool mercury arc. The chief objective was to observe coherently the variety of phenomena reported individually by many investigators and, when possible, to use improved techniques to provide new data and consistent interpretations. Those studies which were a repetition of the work of other investigators gave in general excellent agreement; some of the results of this work point to correct solutions where contradictory results have existed in the past.

Quenching of the 6^3P_1 state.—The only feasible mechanism for the formation of the ion required that the 6^3P_1 state be quenched in a manner very similar to that described by the Stern-Volmer formula. The quenching at 55 mm of N_2 was approximately 66%.

The 6^3P_0 atom.—A method for measuring relative populations of the 6^3P_0 atom is described, which uses the self-absorption of the 4047 A line. This technique was checked and used in the absorption range up to 24% and at pressures as high as 55 mm. The population varies as the first power of the 2537 A intensity. Some of the work of other investigators which disagrees with this relation may be explained by the theory of Gaviola, which concludes that at high intensities the metastable population is proportional to the square root of the intensity. The population was also measured as a function of nitrogen pressure, and the results agreed satisfactorily with the theory of Klumb and Pringsheim. The decay constant of the metastable atoms was studied as a function of pressure, and the results were consistent with the theory of Samson.

The 4850 A continuum.—Additional evidence is presented to show that the molecule responsible for the continuum around 4850 A is formed according to the mechanism

$$Hg(6^3P_1)+N_2 \rightarrow Hg(6^3P_0)+N_2'$$
$$Hg(6^3P_0)+Hg(6^1S_0)+N_2 \rightarrow Hg_2'+N_2.$$

This conclusion is based upon the observations that the intensity of the continuum varies as the first power of the 2537 A radiation, and secondly, that this mechanism explains the manner in which the intensity varies with nitrogen pressure. The mean lifetime for this emission in decay is several times longer than the decay time of the 6^3P_0 atoms or of the rate of ion production.

The resonance cell ion.—The ion originally discovered by Steubing was found in this experiment to exhibit the characteristics reported by numerous observers. The rate at which this ion was produced is dependent on the second power of the 2537 A intensity. Despite the insistence of Biberman that the ion is produced at the surface of the electrodes, the present work definitely confirms Rouse and Giddings' conclusions that the ions are formed in the volume of the gas. It is concluded that the probable mechanism for the formation of the ion is

$$Hg(6^3P_1)+Hg(6^3P_0)+N_2 \rightarrow Hg_2''+N_2$$
$$Hg_2'' \rightarrow Hg_2^+ + e.$$

This mechanism requires a three-body collision of a 6^3P_1 atom and a 6^3P_0 atom with a N_2 molecule to form a highly excited mercury molecule; this molecule is then ionized in a process that is not completely understood. The observation that the decay time of the rate of ion formation is greater than the decay time of the 6^3P_0 state requires that the molecule be formed in an intermediate step prior to its ionization.

The dependence of this ion upon the population of 6^3P_0 atoms was directly demonstrated when it was shown that the ionization rate was decreased when the metastable population was selectively reduced by absorption of 4047 A radiation. The 4047 A intensity was only sufficient to demonstrate the effect; much greater intensities would be needed to distinguish between molecule formation by 6^3P_0 and 6^3P_0 and by 6^3P_1 and 6^3P_0.

From the study of the number of 6^3P_0 atoms required for the formation of the highest ion currents measured, it is concluded that certainly 1 out of every 700 mercury atoms was in the 6^3P_0 state, and that reasonable arguments could place the true value one or two orders of magnitude higher. This offers support to theories which have predicted such high concentrations.

Multiple Isotropic Scattering*

PAUL I. RICHARDS
Technical Operations, Inc., Arlington, Massachusetts
(Received April 21, 1955)

A formalism is derived for approximately solving problems in the transport of radiation or particles by isotropic scattering with absorption. Although the present theory is very similar to diffusion theory, comparison with rigorous solutions where available shows that the results are more accurate than those of diffusion theory especially in highly absorbing media and in regions close to sources. Moreover certain ambiguities and difficulties in extending diffusion theory are eliminated. The derivation proceeds directly from rigorous transport equations and is based on a series expansion for integrals of the Helmholtz type.

1. INTRODUCTION

1.1 Need for a Simple Theory

THE transport or "diffusion" of various radiations and particles through scattering media is governed by the classical Boltzmann transport equation originally derived in the kinetic theory of gases. Unfortunately, rigorous solution of this equation involves mathematical difficulties of a high order. Only recently[1,2] has the simplest problem been rigorously solved, namely the distribution due to an isotropic point source in an infinite isotropically scattering medium. Recent years have also seen numerical solutions[3,4] for a class of problems involving uniformly illuminated plane slabs.

In the vast majority of problems, however, recourse must generally be made to diffusion theory.[5,6] The latter is simple enough to allow solution of a much wider range of problems but suffers several unfortunate drawbacks. Since the concept of diffusion in response to a concentration-gradient is itself an approximation, the coefficient of diffusivity is poorly defined and several values have been proposed.[1,5,6] Secondly, simple diffusion theory cannot be realistic in regions where the motion is highly organized in direction as is the case near a source or in a highly absorbing medium. The results therefore cannot be trusted in such regions where the "direct beam" may be dominant. Finally, it is difficult to establish realistic boundary conditions where the scattering medium gives way to vacuum or a pure absorber. This is espically true if the boundary is directly illuminated by an external source.

In spite of these disadvantages, diffusion theory has been widely and usefully applied to problems where other theoretical methods are intractable.

1.2 Scope of Present Study

It is the aim of this paper to develop a formalism for isotropic scattering as simple and as widely useful as diffusion theory, but possessing the following advantages:

(a) unequivocal values for constants;
(b) definite boundary conditions even under external illumination;
(c) more accurate results particularly in regions near a source and in media which are highly absorbing.

Basic relations are set up in paragraph 2, the approximate formalism is developed in paragraph 3 and a comparison with other theories is given in paragraph 4.

2. EXACT RELATIONS

2.1 Limitations and Notation

We consider throughout the following only media which can absorb and scatter isotropically. Although the formalism can, like the diffusion theory, be applied to media which scatter anisotropically, the reduction to an equivalent isotropic scattering is outside the scope of the present discussion. For definiteness in terminology, we shall speak of light scattering, although of course the theory covers neutrons and other radiations as well.

Only steady-state distributions will be considered; time dependent problems are outside the scope of this study but can be reduced (reference 2, p. 47) to steady state problems.

Vacuum regions will be replaced by absorbing media in the formal development because absorbers are easier to treat in a unified manner and the two have the same effect on photon distribution within other media; neither returns photons.[7]

With these conventions it will then be assumed that

* This work was sponsored by Air Force Cambridge Research Center, Air Research and Development Command.

[1] C. C. Grosjean, Verhandel Koninke Vlaam. Acad. Wetenschap. Belgie, **13**, No. 36 (1951); also Physica **19**, 29 (1953).

[2] Case, deHoffmann, and Placzek *Introduction to the Theory of Neutron Diffusion*. (U. S. Government Printing Office, Washington, D. C., 1953), Vol. 1, Chap. 4.

[3] S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, Oxford, 1950).

[4] Chandrasekhar, Elbert, and Franklin, Astrophys. J. **115**, 244–268, 269–278 (1952).

[5] S. Glasstone and M. G. Edlund, *Elements of Nuclear Reactor Theory* (D. Van Nostrand Company, Inc., New York, 1952).

[6] R. F. Christy in *Lecture Series in Nuclear Physics* (U. S. Government Printing Office, Washington, D. C., 1947), pp. 115–132.

[7] Obvious modifications in treatment are required if photons may stream through vacuum from one scattering region to another. These modifications will not be explicitly considered here. Although the difficulties so introduced can be great, the present theory is no more vulnerable in this respect than others.

the (total) mean free path λ is constant throughout all space. To simplify expressions we will often employ length units chosen so that $\lambda=1$.

The "albedo" of scattering, $\omega(\mathbf{r})$, however, will be allowed to vary arbitrarily with position. (ω = scattering cross section/total cross section $=\lambda/\lambda_{\text{scat}}=1-$probability of absorption per encounter.) In particular $\omega=0$ in vacuum regions.

Other notations which will be used below are as follows:

v = velocity of photons,
ρ = density = number of photons/cm^3,
$\mathbf{j}$ = current = *net* vector flux of photons/cm^2 sec,
q = number of photons emitted by sources/cm^3 sec,
ρ_d = density of unscattered photons,
$\mathbf{j}_d$ = *net* flux of unscattered photons.

The latter two quantities refer to "direct beam" contributions. Note that where several such beams cross at a point, $\mathbf{j}_d$ represents their vector sum whereas ρ_d represents the sum of their magnitudes, similarly $q(\mathbf{r})$ is a scalar sum of photons emitted per unit volume per unit time. Note that the sources need not be isotropic.

2.2 Basic Equations

As shown in standard references[2,3] the basic equation for transport by isotropic scattering may be written

$$\rho(\mathbf{r})=\rho_d(\mathbf{r})+\int_{-\infty}^{\infty}\int\int\rho(\mathbf{r}')\omega(\mathbf{r}')\frac{e^{-|\mathbf{r}-\mathbf{r}'|/\lambda}}{4\pi|\mathbf{r}-\mathbf{r}'|^2\lambda}d\tau'. \quad (1)$$

This relation is easily derived by considering the photon density at $\mathbf{r}$ as made up of (a) direct beam contribution plus (b) photons scattered at $\mathbf{r}'$ and traveling without further scattering to position $\mathbf{r}$. The integral in (1) is to be taken over all space but $\omega(\mathbf{r}')$ is understood to vanish by the convention adopted above in regions where no scattering can occur.

It should be noted that Eq. (1) together with given $\rho_d(\mathbf{r})$ and $\omega(\mathbf{r})$ completely determine the solution of the entire problem. Once the density is known [Eq. (1) has a unique solution[2] for given $\rho_d(\mathbf{r}), \omega(\mathbf{r})$], the current and other quantities are completely determined. There is no need to use the current in setting boundary conditions on (1).

Nevertheless, since the current is often of interest in the results, we mention here some relations which will be used below.

$$\nabla\cdot\mathbf{j}=q(\mathbf{r})-[1-\omega(\mathbf{r})]v\rho(\mathbf{r})/\lambda, \quad (2)$$

$$\nabla\cdot\mathbf{j}_d=q(\mathbf{r})-v\rho_d(\mathbf{r})/\lambda, \quad (3)$$

$$\nabla\times\mathbf{j}=0=\nabla\times\mathbf{j}_d. \quad (4)$$

The first two relations are essentially conservation equations. The first states that photons are gained from the sources and lost only by absorbing collisions. Equation (3) states that *unscattered* photons are supplied by the sources and lost be *either* absorbing or scattering encounters.

Equation (4) may be readily proved by setting up integral expressions for $\mathbf{j}$ and $\mathbf{j}_d$ in terms of ρ, ρ_d and q by the method suggested under (1). Taking the curl of these expressions then establishes (4). Note that (4) implies that $\mathbf{j}$ and $\mathbf{j}_d$ are completely determined by the quantities on the right-hand sides of (2) and (3).

3. SIMPLIFIED THEORY

3.1 Derivation

Since λ has been forced to be constant everywhere we may apply to Eq. (1) the expansion (A.1) derived in the Appendix. Upon evaluating the appropriate integrals defined by (A.2), the result is

$$\rho=\rho_d+\sum_0^{\infty}(\lambda^2\nabla^2)^n(\omega\rho)/(2n+1), \quad (5)$$

where ∇^{2n} represents an n-fold application of the Laplacian operator.

If terms with $n>1$ are omitted in (5), the result is the standard diffusion equation but with a source term equal to $\rho_d v/\lambda$. However, since the coefficients in (5) fall off only as n^{-1}, neglect of higher order terms is probably not very accurate.

To obtain a more rapidly converging series, operate on both sides of (5) with the operator $(\lambda^2\nabla^2-1)$. Rearranging terms, the result is

$$\lambda^2\nabla^2[(3-2\omega)\rho/3]-(1-\omega)\rho = (\lambda^2\nabla^2-1)\rho_d+2\sum_2^{\infty}(\lambda^2\nabla^2)^n(\omega\rho)/(4n^2-1). \quad (6)$$

In Eq. (6), the coefficients of high-order terms now fall off very rapidly and we shall make the approximation of dropping the final sum. In this way it is natural to expect that the approximate results so obtained will be more accurate than diffusion theory. Later comparisons will bear out this expectation.

The formalism is considerably simplified if we introduce an auxiliary function $P(\mathbf{r})$ defined by the following equation:

$$\rho(\mathbf{r})=3[P(\mathbf{r})+\rho_d(\mathbf{r})]/[3-2\omega(\mathbf{r})]. \quad (7)$$

Making this substitution in (6) and dropping the infinite sum, one obtains

$$\lambda^2\nabla^2P(\mathbf{r})-K^2P(\mathbf{r})=-(1-K^2)\rho_d(\mathbf{r}), \quad (8)$$

where

$$K(\mathbf{r})=[3(1-\omega)/(3-2\omega)]^{\frac{1}{2}}. \quad (9)$$

Equation (8) with the definitions (7) and (9) forms the basic equation of the present formalism. The functions $\omega(\mathbf{r})$, the albedo for a single encounter, and $\rho_d(\mathbf{r})$, the density due to direct beams, are presumed to be given data for each specific problem.

Before discussing boundary conditions on $P(\mathbf{r})$, it is convenient to derive an expression for the net photon current. From Eq. (4) it follows that both $\mathbf{j}$ and $\mathbf{j}_d$ can be represented as gradients of scalar functions. Consequently we may also set

$$\mathbf{j}-\mathbf{j}_d=\lambda v\nabla f. \tag{10}$$

Subtracting (3) from (2) and making the substitution (10), one obtains

$$\lambda^2\nabla^2 f=\rho_d-(1-\omega)\rho. \tag{11}$$

It then follows from (7), (8), and (9) that

$$\nabla^2(f+P)=0. \tag{12}$$

Both P and f (aside from an undetermined constant) must vanish at infinity since ρ and $\mathbf{j}$ must vanish there. If we envisage any geometrical discontinuities in scattering media (i.e., discontinuities in K and ω) as replaced by very sharp but nevertheless continuous transitions, then on physical grounds all functions must be continuous with their first derivatives and it follows from (12) that f may be taken equal to $(-P)$. Equation (10) then becomes

$$\mathbf{j}=\mathbf{j}_d-\lambda v\nabla P. \tag{13}$$

This relation suffices to determine the net current when (8) has been solved for P.

Note that the derivation has involved no further approximations beyond the use of Eq. (8). Hence $\mathbf{j}$ as determined from (13) will satisfy rigorously the conservation and curl relations (2), (3), and (4) but the density appearing in these will be the approximate density as determined from (7) and (8). In short, Eq. (13) will yield precisely the current implied by the approximate ρ.

3.2 Boundary Conditions

In solving actual problems, it is convenient to consider changes in physical properties between various media as discontinuous. It then becomes necessary to specify the behavior of the fundamental auxiliary function P at such discontinuities.

Since photons do not accumulate at boundaries, it follows from (13) that the normal component of ∇P is continuous. This result may also be derived from (8) by integrating over a "pill box" volume which straddles the interface and letting the height of the box approach zero.

To derive the second boundary condition, it is convenient to consider the limiting case of a continuous transition between media as the distance through the transition is made to approach zero. Assume first that P becomes discontinuous in the limit. Then as the distance of transition is made small, the normal component of the current $\mathbf{j}$ will, by (13) attain an arbitrarily large value within the transition while remaining finite on either side of the boundary. This situation can only represent a "double layer" of sources and sinks distributed over the boundary, and must be ruled out of any physical problem not involving such double layers. Hence the remaining boundary condition is that P itself is continuous.

It may be noted that ρ as given by (7) will then be discontinuous at boundaries. This is not in general true of a rigorous solution,[2] but the latter often has a sudden drop very near an absorbing boundary, and the discontinuity given by the present formalism is presumably an approximation to this sudden drop.

It is of interest to note that the simple formalism presented here also automatically gives an extrapolated endpoint.[6] Specifically, if we assume a plane boundary located in a region where $\rho_d=0$ and separating a vacuum or an absorbing medium from a medium with finite ω, then the solution of (8) on the absorbing side is of the form $\exp(-|Z|/\lambda)$ where Z is distance measured normally from the boundary. By continuity of P and its normal derivative, the value and derivative of ρ will then have the ratio λ just inside the interface, and a linear extrapolation of ρ into the absorbing medium will then vanish at a distance λ from the boundary. Thus the formalism gives an extrapolated end point of λ rather than the rigorous value,[2,6] 0.7104λ. However, the present formalism will automatically alter this value to take account of various specific local conditions. There is no need, as in diffusion theory, to use separate arguments to establish its appropriateness in any given situation.

3.3 Summary

Here we collect the above results in a form which would be used to solve a specific problem. For simplicity it is assumed that there are no reflecting boundaries and that the geometry is such that photons cannot stream across a vacuum from one scattering medium to enter another. While the complications thus excluded can be very great, they are not peculiar to the present formalism.

The function, $\omega(\mathbf{r})$ = the probability that an encounter at $\mathbf{r}$ results in (isotropic) scattering rather than absorption, is specified by the properties of the various scattering media. It is assumed that the total mean path λ is constant in all media.

The function, $\rho_d(\mathbf{r})$ = density of unscattered photons, is readily computed from the given distribution of sources. Note that the sources need not themselves be isotropic and that they can be located either inside scattering media or in vacuum. Values of $\rho_d(\mathbf{r})$ will be needed only in regions where $\omega\neq 0$.

After $\rho_d(\mathbf{r})$ has been computed, vacuum regions are replaced by complete absorbers ($\omega=0$). The function $K(\mathbf{r})$ is then computed from

$$K=[(1-\omega)/(1-2\omega/3)]^{\frac{1}{2}}. \tag{14}$$

This completes the setup of the problem.

Solution of the problem is then effected by solving the following equation for the auxiliary function $P(\mathbf{r})$

$$\lambda^2\nabla^2 P-K^2P=-(1-K^2)\rho_d, \tag{15}$$

TABLE I. Comparison of $K(\omega)$ for various theories.

ω	$K(\omega)$ By (14)	$K(\omega)$ Rigorous[a]	$K(\omega)$ Diffusion theory
0	1	1	1.73
0.1	0.982	1.0000	1.64
0.2	0.960	0.9999	1.55
0.3	0.935	0.997	1.45
0.4	0.904	0.986	1.34
0.5	0.867	0.957	1.22
0.6	0.817	0.907	1.09
0.7	0.750	0.829	0.949
0.8	0.655	0.710	0.775
0.9	0.500	0.525	0.548
0.94	0.402	0.414	0.425
0.98	0.240	0.243	0.245
0.99	0.172	0.1725	0.173
1	0	0	0

[a] See reference 2.

subject to the boundary conditions

$$P \to 0 \text{ as } \mathbf{r} \to \infty$$

$$P \text{ and } \mathbf{n}\cdot\nabla P \text{ continuous at interfaces.} \qquad (16)$$

The density ρ of photons and the *net* flux $\mathbf{j}$ are then given by

$$\rho = (P+\rho_d)/(1-2\omega/3), \qquad (17)$$

$$\mathbf{j} = \mathbf{j}_d - \lambda v \nabla P. \qquad (18)$$

In (18) $\mathbf{j}_d$ is the *net* flux of unscattered photons, which is easily computed from the given source distribution (with, of course, vacuum not replaced by absorbers).

Equations (17) and (18) give ρ and $\mathbf{j}$ assuming vacuum to act as a perfect absorber. Under the limitations stated above, they are thus correct (subject to the approximation of the whole theory) within scattering media, but of course do not represent the actual situation in vacuum regions. If needed, the latter must be separately computed from the former.

The value (18) for $\mathbf{j}$ will satisfy rigorously the conservation relations implied by the approximate value of $\rho(\mathbf{r})$.

It will be seen that the effort of solving a problem is about the same as that required by diffusion theory. Nevertheless, the results are shown below to be more accurate in cases where comparison with rigorous solution is possible.

TABLE II. Values of $4\pi r^2 \rho v/S$ from various theories. Point source in an infinite medium, $\lambda = 1$.

ω	r	By (19)	Rigorous[a]	Simple diffusion theory	Diffusion theory with rigorous K
	0	2.5	1	0	0
	0.5	1.96	1.52	1.14	1.15
	1	1.84	1.80	1.73	1.77
	1.5	1.77	1.90	1.97	2.05
0.9	2	1.68	1.87	2.01	2.10
	3	1.43	1.60	1.59	1.86
	4	1.13	1.24	1.33	1.47
	5	0.85	0.92	0.97	1.08
	6	0.62	0.65	0.66	0.77
	7	0.44	0.45	0.46	0.53
	0	1.25	1	0	0
	0.5	0.85	0.77	0.73	0.91
	1	0.58	0.54	0.70	1.11
	1.5	0.40	0.37	0.51	1.01
0.3	2	0.28	0.24	0.33	0.82
	3	0.13	0.10	0.12	0.45
	4	0.060	0.043	0.036	0.22
	5	0.028	0.017	0.010	0.10
	6	0.012	0.007	0.003	0.045
	7	0.005	0.003	0.0008	0.019

[a] See reference 2.

4. COMPARISON WITH OTHER THEORIES

4.1 Point Source in Infinite Medium

The present formalism is easily applied to the problem of an isotropic point source in an infinite homogeneous (isotropically) scattering and absorbing medium. In this case we have immediately $\rho_d = S\exp(-r)/4\pi r^2 v$ where S is the source strength in photons/sec and the unit of length has been chosen to make $\lambda = 1$. The parameter K is a constant and Eq. (15) is readily solved by the method of Green's function. The final result is

$$\rho = [S/4\pi v(1-2\omega/3)]\{e^{-r}/r^2 + [(1-K^2)/2Kr] \times [e^{-Kr}2\tanh^{-1}K + e^{Kr}E_1((1+K)r) - e^{-Kr}E_1((1-K)r)]\}, \qquad (19)$$

where

$$E_1(X) = -\mathrm{Ei}(-X) = \int_X^\infty (e^{-u}/u)du. \qquad (20)$$

It may be noted that when ω vanishes (19) becomes the rigorous solution for a perfectly absorbing medium. It may also be shown that for all ω the total number of photons in the distribution (19) has the rigorous[2] value $S/(1-\omega)v$.

In comparing (19) with the corresponding result for diffusion theory, we run across one of the disquieting features of the latter. Simple diffusion theory gives the value $[3(1-\omega)]^{\frac{1}{2}}$ for K. This, of course, is greatly in error for small ω as may be seen in Table I where it also is shown that the value (14) for K is much closer to the rigorous[2] value. In view of this discrepancy, various corrected expressions are often used[4] in diffusion theory. Whatever the value of K employed, diffusion theory with a δ-function source term gives the result $\rho = (3S/4\pi v)(e^{-Kr}/r)$.

Now this expression gives the correct total number of photons only if the *less* accurate value of K is used. Moreover, a similar feature arises in comparing detailed values of ρ. In Table II are shown values of ρ given by various theories for $\omega = 0.9$ and $\omega = 0.3$. For values of ω near unity, there seems to be little choice between the various approximate theories (in this example at least). For $\omega = 0.3$, we note again that the "corrected" diffusion theory is rather poor whereas the simple diffusion theory is surprisingly good in spite of the fact that K is 45% too

TABLE III. Transmission of nonabsorbing plane slab. ($\lambda=1, \omega=1$)

$\cos\theta$	t	Transmission: This paper	Transmission: Rigorous[a]	Transmission: Diffusion theory	$\cos\theta$	t	Transmission: This paper	Transmission: Rigorous[a]	Transmission: Diffusion theory
	0.05	0.975	0.975	0.98		0.05	0.683	0.680	0.632
	0.1	0.953	0.953	0.95		0.1	0.561	0.549	0.432
1	0.25	0.889	0.890	0.885	0.05	0.25	0.469	0.430	0.200
	0.5	0.802	0.797	0.786		0.5	0.420	0.356	0.100
	1	0.667	0.660	0.632		1	0.350	0.275	0.050
	0.05	0.953	0.954	0.950		0.05	0.488	0.456	0
	0.1	0.910	0.910	0.905		0.1	0.477	0.429	0
0.5	0.25	0.802	0.800	0.786	0	0.25	0.445	0.372	0
	0.5	0.674	0.666	0.632		0.5	0.400	0.313	0
	1	0.522	0.502	0.432		1	0.333	0.242	0
	0.05	0.890	0.890	0.885					
	0.1	0.803	0.801	0.786					
0.2	0.25	0.637	0.627	0.570					
	0.5	0.507	0.480	0.368					
	1	0.402	0.354	0.200					

[a] See reference 4.

large. At very great distances r, however, the reverse must be true since the exponential function will eventually dominate the behavior. Thus, in all, it appears that the present formalism, with no additional corrections, produces a single set of formulas which are much better approximations over large ranges of the parameters.

4.2 Plane Slab Problems

Consider a homogeneous plane slab of thickness t, bounded on both sides by vacuum. Let the slab be uniformly illuminated by parallel radiation making an angle θ with the normal. This problem also has been rigorously solved[3,4] for isotropic scattering.

In Table III, we show the transmission (=power transmitted/power incident) of a nonabsorbing slab as obtained from the various theories. It will be seen that the present formalism in all cases gives results which are better than those of diffusion theory and that in particular it predicts about the right transmission for grazing incidence ($\theta=\pi/2$) while diffusion theory gives a limiting value of zero. In Table III moreover we also have another example of the uncertainty entailed in any simple attempt to correct diffusion theory. The values shown are for the boundary condition $\rho=0$; in this example, use of the extrapolated endpoint (equivalent to using $t+1.4\lambda$ as the thickness) would make the diffusion approximation worse.

In Table IV are shown values of the albedo (=power reflected/power incident) for infinite thickness and various ω, θ. Here again it is seen that the present formalism is a better approximation over the full range of the parameters. Again, the results shown for diffusion theory are those of the simple version; use of diffusion formulas with the rigorous value of K makes the albedo approximation worse although the detailed variation of ρ is presumably better.

TABLE IV. Albedo of half-space.

ω	$\cos\theta$	Albedo: This paper	Albedo: Rigorous[a]	Albedo: Diffusion theory	ω	$\cos\theta$	Albedo: This paper	Albedo: Rigorous[a]	Albedo: Diffusion theory
	0	0	0	1		0	0.250	0.452	1
0	0.5	0	0	0.536	0.7	0.5	0.182	0.278	0.678
	1	0	0	0.366		1	0.143	0.208	0.513
	0	0.018	0.052	1		0	0.500	0.694	1
0.1	0.5	0.012	0.024	0.550	0.9	0.5	0.400	0.508	0.785
	1	0.009	0.016	0.378		1	0.333	0.415	0.645
	0	0.065	0.163	1		0	0.719	0.776	1
0.3	0.5	0.044	0.082	0.580	0.95	0.5	0.631	0.626	0.837
	1	0.034	0.058	0.408		1	0.561	0.536	0.720
	0	0.133	0.293	1		0	1	1	1
0.5	0.5	0.093	0.161	0.620	1	0.5	1	1	1
	1	0.071	0.115	0.448		1	1	1	1

[a] See reference 3.

4.3 Conclusions

In summary, an approximate formalism has been derived directly from the rigorous isotropic transport equations. Although the labor involved in the solution of specific problems is about the same as that required in diffusion theory, the results appear to be a better approximation particularly for absorbing media. The results are, of course, approximate but if further improvement is required, it is clear from the derivation itself what further steps should be taken to increase the accuracy. This is in contrast to diffusion theory where, as noted above, apparently reasonable attempts to improve accuracy may in fact have the opposite effect. It appears that the present formalism should be widely applicable to problems too complex to be handled by rigorous methods, and we expect to make a number of such applications in the near future.

APPENDIX

The purpose of this appendix is to indicate the derivation of the following expansion.

$$(4\pi)^{-1}\int_{-\infty}^{\infty}\int\int F(\mathbf{r}')K(|\mathbf{r}'-\mathbf{r}|)d\tau' = \sum_{0}^{\infty} a_n \nabla^{2n}F(\mathbf{r})/(2n+1)! \quad \text{(A.1)}$$

where

$$a_n = \int_0^\infty u^{2n+2}K(u)du. \quad \text{(A.2)}$$

The notation ∇^{2n} denotes an n-fold repeated application of the Laplacian operator. No detailed examination of restrictions will be made, but obviously necessary conditions are the convergence of (A.2) for all n and the validity of the Taylor series expansions for F.

To evaluate the integral in (A.1), change the integration variable to $\mathbf{R}=\mathbf{r}'-\mathbf{r}$ and expand $F(\mathbf{R}+\mathbf{r})$ in a three-dimensional Taylor series about $\mathbf{r}$.

$$\int_{-\infty}^{\infty}\int\int [e^{\mathbf{R}\cdot\nabla}F(\mathbf{r})]K(R)d\tau_R/4\pi,$$

where the exponential operator is defined by its power series. The mth term in this series is itself a sum of 3^m terms which arise from the scalar product in the exponent. Collecting like powers of these ultimate terms, the mth term of the main series can be written

$$\frac{1}{m!}\int_0^\infty K(R)R^{m+2}dR\cdot\sum_{(i+j+k=m)}\binom{m}{i}\binom{m-i}{j} \cdot\left\{\left(\frac{\partial}{\partial x}\right)^i\left(\frac{\partial}{\partial y}\right)^j\left(\frac{\partial}{\partial z}\right)^k F(r)\right\}\oint\oint\frac{X^iY^jZ^k}{4\pi R^m}d\Omega_R. \quad \text{(A.3)}$$

Now by spherical symmetry, the surface integrals in this expression will vanish unless i, j, and k are all even. Consequently, the entire expression (A.3) vanishes unless m is even. The surface integrals in the remaining cases can be evaluated from Dirichlet's integral,[8] a special case of which is (for i, j, k even)

$$\int_{r\le a}\int\int x^iy^jz^kd\tau = a^{m+3}\Gamma\left(\frac{i+1}{2}\right)\Gamma\left(\frac{j+1}{2}\right)\Gamma\left(\frac{k+1}{2}\right)\Big/\Gamma\left(\frac{m+5}{2}\right).$$

Differentiating this with respect to a and then setting $a=1$, the integral required in (A.3) is obtained. Substituting the result in (A.3), combining numerical factors, and recalling that i, j, and k are all even, it is found that (A.3) reduces to the term $n=m/2$ in (A.1) and the result follows.

It is perhaps of interest to note that the case $K(x) = -e^{-x}/x$ yields a series (A.1) formally equivalent to the operator $(\nabla^2-1)^{-1}$ thus indicating the well-known fact that this particular K is the Green's function of the operator (∇^2-1).

[8] See, for example, R. C. Tolman, *Principles of Statistical Mechanics* (Oxford University Press, Oxford, 1938), p. 656.

Lifetime of Electrons in p-Type Silicon

G. Bemski
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received June 27, 1955)

The dependence of lifetime of electrons on temperature in p-type silicon can be explained on the basis of the Shockley-Read-Hall theory by assuming a level for the recombination centers at 0.2 ev from the valence band. The variation of lifetime with injected carrier density can also be explained by the same model.

1. INTRODUCTION

MEASUREMENTS described in this paper of minority carrier lifetime in the vicinity of a silicon p-n junction show that lifetime is a function of both minority carrier density and temperature. These experiments provide an indication of the processes involved in the recombination of holes and electrons. The results are consistent with the Shockley-Read-Hall theory of recombination centers.[1,2]

2. THEORY

The Shockley-Read-Hall theory of electron-hole recombination in semiconductors explains why lifetimes observed in germanium and silicon are several orders of magnitude shorter than those predicted from direcs electron hole recombination. The model employed it shown in Fig. 1. It assumes that recombination takes place through the agency of imperfections of some sort in the crystal.

The recombination process can be characterized by the concentration of centers N_t, the energy level of centers E_t, the thermal densities of electrons and holes n_0 and p_0, and the capture constants for holes and electrons $C_p=1/\tau_{p0}$ and $C_n=1/\tau_{n0}$. τ_{p0} is the lifetime for holes injected into high-conductivity n-type specimen, τ_{n0} is similarly the lifetime of electrons in high-conductivity p-type specimen. C_n and C_p are assumed constant in the temperature range investigated.

For small values of injected carrier density δn the theory gives for the lifetime τ_0:

$$\tau_0=\frac{\frac{1}{C_p}(n_0+n_1)+\frac{1}{C_n}(p_0+p_1)}{n_0+p_0}, \tag{1}$$

where

$$n_1=N_c e^{(E_T-E_c)/kT}$$

and

$$p_1=N_v e^{(E_v-E_T)/kT}$$

represent the numbers of electrons and holes in the conduction and valence bands for the case in which Fermi level falls at E_t, and N_c and N_v are the effective densities of states in the conduction and valence bands, respectively. For large values of injected carrier densities δn Shockley and Read obtain:

$$\tau=\tau_0\frac{\left(1+\frac{\delta n(\tau_{p0}+\tau_{n0})}{\tau_{p0}(n_0+n_1)+\tau_{n0}(p_0+p_1)}\right)}{1+\frac{\delta n}{n_0+p_0}}. \tag{2}$$

Equations (1) and (2) provide means for determining the energy level of the recombination centers and the ratio of capture cross sections for electrons and holes assuming that all the centers lie at a single energy level within the energy gap.

3. EXPERIMENT

The junction recovery time[3] method of measuring lifetime was used in our experiments. This type of measurement yields results depending upon the carrier density within one diffusion length from the p-n junction. By using p-n grown junctions of large cross section one may observe lifetime very close to the body lifetime even in the presence of high surface recombination velocity. Each specimen used was cut from a p-n grown junction crystal and etched in CP-4 to a size of 0.100 in.×0.100 in.×0.500 in. The resistivity of the p side was 50 to 100 times higher than that of the n side. This disparity insured predominating injection of electrons from the low resistivity to the high-resistivity side. The electron lifetime was then measured on the high-resistivity side of the junction.

The effects of the space-charge region of the junction have been neglected in view of the large ratio of injected carrier density to the capacitance current density.[3] Some of the crystals were rejected when the oscilloscope trace indicated more than one time con-

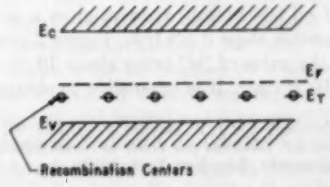


FIG. 1. Model for the recombination of electrons and holes.

[1] W. Shockley and W. T. Read, Phys. Rev. **87**, 835 (1952).
[2] R. N. Hall, Phys. Rev. **87**, 387 (1952).
[3] B. Lax and S. Neustadter, J. Appl. Phys. **25**, 1148 (1954).

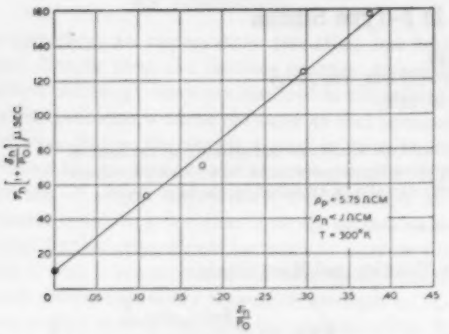


FIG. 2. Lifetime in p-type silicon vs injected electron density.

stant. This could have been due to defective junctions or to high trap densities near the junctions.[4] There was no indication of traps in the specimens retained for our experiment; the lifetimes were not sensitive to dc light and were found to decrease with the cross sections of the samples according to theory.[5]

4. RESULTS

Typical measurements of lifetime as a function of injected carrier density are shown in Fig. 2. Similar results have been obtained on samples from different crystals with resistivities of the p sides varying between 1 and 15 ohm-cm. The data are plotted as $\tau_n(1+\delta_n/p_0)$ vs $\delta n/p_0$ to facilitate comparison with theory (next section). The points have been taken at a temperature of 300°K. Figure 3 shows the points obtained in measuring lifetime vs inverse absolute temperature.

5. DISCUSSION

To interpret the data of Fig. 2 one can use a simplified form of Eq. (2). In view of the result described farther on and indicating that the recombination centers are located in the lower half of the energy gap, we assume that p_1 is larger than n_0 and n_1; τ is then given by:

$$\tau_n=\frac{\tau_0\left(1+\dfrac{\delta n(\tau_{p0}+\tau_{n0})}{\tau_{n0}(p_0+p_1)}\right)}{1+\dfrac{\delta n}{p_0}}=\frac{\tau_0(1+a\delta n)}{1+\dfrac{\delta n}{p_0}}.$$

A plot of $\tau_n[1+(\delta n/p_0)]$ vs $\delta n/p_0$ gives a straight line with a positive slope if $a>1/p_0$. Figure 2 confirms this relation, the value of "a" being about 10^{-14} cm^{-3} while $1/p_0=4.10^{-16}$ cm^{-3}. It is impossible to obtain the cross section for electron capture from this expression as neither τ_{n0} nor τ_{p0} is known. It can be estimated, however, that the ratio $\tau_{p0}/\tau_{n0}=C_n/C_p$ is considerably larger than unity. This would indicate that the cross section of the empty traps for an electron capture is higher than that of occupied traps for capture of a hole. It is interesting to note that Burton[6] has found, in the case of copper and nickel recombination centers in germanium, a result opposite to ours.

The experimental points in Fig. 3 can be fitted by an empirical expression of the form:

$$\tau_0=\tau_{n0}+Ce^{-(E_t/kT)}.$$

If the recombination centers lie in upper half of the gap, Eq. (1) would give

$$\tau_0=\tau_{n0}+\tau_{p0}(n_1/p_0),$$

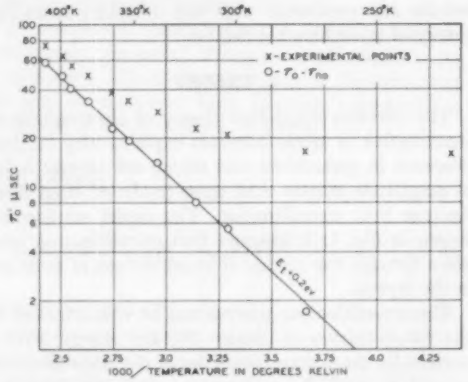


FIG. 3. Lifetime in 20Ω-cm p-type silicon vs $1000/T$. The slope of $\tau_0-\tau_{n0}$ gives the level of recombination centers within the energy gap.

and since $\tau_{p0}\gg\tau_{n0}$ and $n_1\geq p_0$,

$$\tau_0\approx\tau_{p0}(n_1/p_0).$$

For centers in lower half of the gap ($p_1\gg n_1$),

$$\tau_0=\tau_{n0}[1+(p_1/p_0)]=\tau_{n0}+Ce(E_v-E_t)/kT,$$

in agreement with the experimental data. Taking $\tau_{n0}=15$ μsec and plotting $\log(\tau_0-\tau_{n0})$ vs $T/1000$ gives a straight line. The slope corresponds to $E_t=0.2$ ev.

ACKNOWLEDGMENTS

The author wishes to thank his colleagues for helpful discussions and, in particular, Dr. R. Shulman for taking the $\delta n\sim 0$ measurement in Fig. 2.

[4] J. A. Hornbeck and J. R. Haynes, Phys. Rev. **97**, 311 (1955).
[5] W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950).
[6] Burton, Hall, Morin, and Severiens, J. Phys. Chem. **57**, 853 (1953).

Effect of Defects on Lattice Vibrations*

ELLIOTT W. MONTROLL AND RENFREY B. POTTS†
University of Maryland, College Park, Maryland

(Received July 13, 1955)

The theory of the effect of localized defects such as impurities, holes, and interstitials on the vibrations of crystal lattices is developed. Although most of the analysis is concerned with one-dimensional chains, the general approach to defects in three-dimensional lattices is outlined through the example of a simple cubic lattice with nearest-neighbor interactions.

Many types of defects cause localized normal modes whose effect dies out rapidly with distance from the defect. Mathematical techniques, which involve the use of Green's functions, are discussed for the theory of these localized modes. The vibrational frequencies of these modes are displaced out of the band of frequencies of a perfect lattice.

The theory of interaction of two defects as a function of their distance of separation is developed for the range of very low temperatures through the calculation of the change of zero-point energy of a lattice as a result of the introduction of a defect pair. Defects attract each other in a monatomic lattice. The attraction between two mass defects in a linear chain is inversely proportional to the cube of their distance of separation.

The effect of a localized defect mode in a simple cubic lattice diminishes as with the distance r as $r^{-1}\exp(-Ar)$.

INTRODUCTION

LOCAL defects such as impurities, holes, interstitials, etc., in crystals make their existence apparent in many striking ways. The impurity levels of semiconductors are the seat of those electrons which can easily make transitions into the conduction band and hence (with the holes which behave in an analogous manner) are to a large degree the dominant influence in the electrical behavior of semiconductors. Impurities are known to change relaxation times in magnetic materials by many orders of magnitude. Transport processes such as electrical and heat conduction are frequently controlled by the scattering of electrons or phonons by defects.

An enormous literature exists which gives a semi-empirical or phenomenological theory of the effects mentioned above as well as of the diffusion of defects. A proper basis for the systematic discussion of the effect of impurities on electrons in solids has been given by Slater and Koster[1] through the use of localized Wannier[2] wave functions. The effect of defects on electrons in a linear chain has also been discussed by Saxon and Hutner.[3] There is considerable similarity between the behavior of electrons in semiconductors, spin waves in magnetic materials and lattice vibrations in crystals in general. Since lattice vibrations are the most amenable of these three phenomena to mathematical analysis, the authors have felt that a detailed investigation of the effect of lattice defects on lattice vibrations might shed some light on the general defect problem.

It is not only in solid state physics that defects are important. Any discussion of the interaction of particles with a field is really one of localized defects in the field. Since the theory of lattice vibrations is mathematically equivalent to that of a discrete quantum field, an analogy might be made between lattice defects and particles coupled to discrete quantum fields. Fortunately a natural fundamental length, the lattice spacing, is sufficient to immunize the theory of lattice vibrations from the divergences which are the curse of continuum quantum fields. It is hoped that some of the results on the discrete model may contribute to the understanding of some of the difficulties in the continuum case. Lattice theories of quantum fields have been discussed recently by Schiff[4] and Rosenstock.[5]

A local defect influences the vibrational spectrum in two ways. With the exception of a small number of frequencies which normally lie near the band edges, the possible frequencies are displaced by amounts of order $O(1/N)$ (N being the total number of lattice points). The exceptional frequencies, which might suffer a considerable displacement, are associated with localized normal modes which die out rapidly with distance from the defect. A typical such mode is that associated with a lattice vacancy. It corresponds to a pulsation (similar to that of an explosion bubble in a fluid) whose influence dies out as $r^{-1}\exp(-Ar)$ at large distances. The large class of slightly displaced modes can be discussed through perturbation theory, but the exceptional localized modes must be handled by a separate exact calculation. The theory of these modes has been briefly outlined by Lax and Smith[6] and is essentially the same as that proposed by Koster and Slater[1] for electrons in solids. Although the approach used in the present paper follows in similar lines, it was first suggested to the

* This work was supported by the Air Research and Development Command of the U. S. Air Force.

† On leave from the University of Adelaide, South Australia.

[1] G. F. Koster and J. C. Slater, Phys. Rev. **95**, 1167 (1954); G. F. Koster, Phys. Rev. **95**, 1436 (1954).

[2] G. H. Wannier, Phys. Rev. **52**, 191 (1937).

[3] D. S. Saxon and R. A. Hutner, Philips Research Repts. **4**, 81 (1949).

[4] L. I. Schiff, Phys. Rev. **92**, 764 (1953).

[5] H. B. Rosenstock, Phys. Rev. **93**, 331 (1954).

[6] M. Lax, Phys. Rev. **94**, 1391 (1954).

authors by a mathematical investigation of discrete potential theory by Duffin.[7]

Every lattice defect has a "self-energy." This corresponds to the difference in bond energy between the imperfect and perfect lattice plus the differences in vibrational zero-point energy, the $\frac{1}{2}h\nu$'s associated with each frequency.

An interesting feature of defects in a monatomic lattice is their clustering tendency. The short-range preference for pairing of vacancies due to the lowering of total bond energies is well known. We shall show that a longer-range attractive force also exists between defects. If the defects are light isotopes imbedded in a matrix of heavier isotopes, this attractive force gives rise to a separation of the isotopic mixture into two phases at very low temperatures. This point has been discussed by perturbation theory by Prigogine, Bingen, and Jeener.[8]

The authors feel that this attraction between defects is a rather general property of locally perturbed fields. In this case it would give rise to the following mechanism: defects might be created by some means of excitation, as they are formed they would cluster together so that when they exist in sufficiently large numbers they might condense in a manner analogous to that in which a gas condenses.

There are three phases of the theory of defects which must be investigated. First, the statics or the energetic considerations such as their self-energies and the interactions between defects; second, the dyamics or motions of defects through the lattice, and finally the statistical mechanics or the effect of thermal excitation on defect formation and interaction. This paper is concerned with the statics. The authors hope to discuss the other two phases later. Stripp and Kirkwood[9] have analyzed the statistical mechanics of holes and impurities in lattices but this work deals only with perturbation theory and neglects the effect of the localized modes which in some situations might be extremely important.

1. DESCRIPTION OF THE MODEL

Our detailed analysis will be carried out for linear chains and simple cubic lattices with nearest neighbor interactions only. Other lattices can be handled in an analogous manner but require somewhat more cumbersome mathematics. We describe our model in this section.

(a) Lattice or Field

Let us consider a simple cubic lattice of N^3 identical particles each of mass M and coupled to its nearest neighbors through both central and noncentral forces. The forces are postulated as harmonic so that the total potential energy of the lattice is

$$\Phi=\Phi_0+\Delta\Phi,$$

where Φ_0 is the potential energy of the lattice at equilibrium, $\Phi_0=3N^3v_0$ (v_0 being the binding energy per pair of particles) and

$$\begin{aligned}\Delta\Phi=&\frac{1}{2a^2}\beta_1\sum_{l,m,n}[(x_{l,m,n}-x_{l+1,m,n})^2\\&+(y_{l,m,n}-y_{l,m+1,n})^2+(z_{l,m,n}-z_{l,m,n+1})^2]\\&+\frac{1}{2a^2}\beta_2\sum_{l,m,n}[(x_{l,m,n}-x_{l,m+1,n})^2\\&+(y_{l,m,n}-y_{l,m,n+1})^2+(z_{l,m,n}-z_{l+1,m,n})^2]\\&+\frac{1}{2a^2}\beta_3\sum_{l,m,n}[(x_{l,m,n}-x_{l,m,n+1})^2\\&+(y_{l,m,n}-y_{l+1,m,n})^2+(z_{l,m,n}-z_{l,m+1,n})^2],\end{aligned}\quad(1.1)$$

when the set of displacements $\{x_{l,m,n},y_{l,m,n},z_{l,m,n}\}$ from equilibrium positions exist at the lattice points $\{l,m,n\}$. The lattice spacing is a. We generally let $\gamma_j=\beta_j/a^2$. The constant γ_1 represents the central force constant while we usually write $\gamma_2=\gamma_3$ for the noncentral force constants. Since a simple cubic lattice with nearest neighbor central force interactions only is unstable to shear we must introduce the noncentral forces to give the system a realistic character.

We shall have occasion to discuss a discrete scalar field whose potential energy can be described as a special case of (1.1). We characterize this field by one parameter $x_{l,m,n}$ associated with each lattice point (l,m,n) and by the potential energy function (1.1) with $\beta_1=\beta_2=\beta_3=\beta$ and $y=z=0$.

We shall first give a classical mechanical discussion of our system and later analyze it quantum mechanically. It is to be recalled that in the case of harmonic operators the quantum mechanical energy levels are $E=\Sigma(n_j+\frac{1}{2})h\nu_j$, where the n_j's are integers and ν_j's the classical frequencies of the normal modes of vibration of the system. Our system of particles is of course equivalent to a triply periodic set of springs and masses. We shall sometimes refer to it as the "field."

Since the motions in the x, y, and z directions are independent of each other in our model we need only analyze the behavior of a lattice with one degree of freedom per lattice point. Newton's equations of motion for such a system are

$$\begin{aligned}M\ddot{x}_{l,m,n}=&(\beta_1/a^2)[x_{l+1,m,n}-2x_{l,m,n}+x_{l-1,m,n}]\\&+(\beta_2/a^2)[x_{l,m+1,n}-2x_{l,m,n}+x_{l,m-1,n}]\\&+(\beta_3/a^2)[x_{l,m,n+1}-2x_{l,m,n}+x_{l,m,n-1}].\end{aligned}\quad(1.2)$$

Similar equations exist for the y's and z's. The equations of motion of our model of a discrete scalar field are the special case of (1.2) with $\beta_1=\beta_2=\beta_3=\beta$.

[7] R. J. Duffin, Duke Math. J. **20**, 233 (1953).
[8] Prigogine, Bingen, and Jeener, Physica **20**, 383 (1954); **20**, 516 (1954).
[9] K. F. Stripp and J. C. Kirkwood, J. Chem. Phys. **22**, 1579 (1954).

We shall choose our system to be composed of $N\times N\times N=N^3$ particles and use periodic boundary conditions

$$x_{l,m,n}=x_{l+N,m,n}=x_{l,m+N,n}=x_{l,m,n+N}, \quad \text{etc.} \tag{1.3}$$

Then the normal modes can be expressed as

$$x_{l,m,n}=u(l,m,n)\exp -i\omega t, \tag{1.4a}$$

where

$$u(l,m,n)=\exp i(\varphi_1 l+\varphi_2 m+\varphi_3 n), \tag{1.4b}$$

and

$$\varphi_j=2\pi a_j/N; \quad a_1,a_2,a_3=1,2,\cdots,N$$

or

$$-\tfrac{1}{2}N \cdots, \tfrac{1}{2}N. \tag{1.4c}$$

The circular frequencies of the normal modes of vibration of our "field" are

$$a^2M\omega^2=2\sum_1^3\beta_j(1-\cos\varphi_j); \tag{1.5a}$$

if $a\to 0$, $\beta\to\infty$, $M\to 0$ in such a way that $\beta/M=\alpha$ = constant and $L=Na$ = length of an edge of our lattice, we find

$$\omega^2\to 4\pi^2L^{-2}(\alpha_1a_1^2+\alpha_2a_2^2+\alpha_3a_3^2) \tag{1.5b}$$

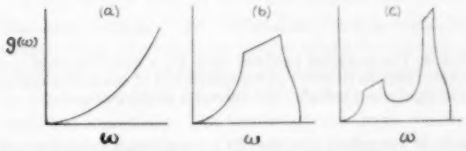


FIG. 1. The distribution function $g(\omega)$ for the number of normal frequencies between ω and $\omega+d\omega$. (a) continuum, (b) $\gamma_1=\gamma_2=\gamma_3$, (c) $\gamma_1>\gamma_2=\gamma_3$.

the square of a typical frequency of a normal mode of the continuum. It is to be recalled that our equations of motion (1.2) reduce to the wave equation (when $\alpha_1=\alpha_2=\alpha_3=\alpha$)

$$\nabla^2u=\alpha^{-2}\partial^2u/\partial t^2. \tag{1.6}$$

An important difference between the continuum and a discrete lattice is that the largest frequency in the discrete lattice is

$$\omega_L^2=4M^{-1}(\gamma_1+\gamma_2+\gamma_3), \tag{1.7}$$

while the set of normal modes of the continuum is unbounded. It is this unboundedness that leads to many of the divergences in field theories.

The distribution function $g(\omega)$ defined so that $g(\omega)d\omega$ is the number of frequencies between ω and $\omega+d\omega$ is quadratic in ω in the continuum [Fig. 1(a)] and has the same character at low frequencies in a lattice, but vanishes at high frequencies [see Figs. 1(b) and 1(c)].

The distribution function of ω^2 is found from its characteristic function

$$f(v)=E(\exp iv\omega^2)$$

$$=\frac{1}{N^3}\sum_{a_1,a_2,a_3=1}^{N}\exp\{2ivM^{-1}\sum_1^3\gamma_j(1-\cos\varphi_j)\}, \tag{1.8}$$

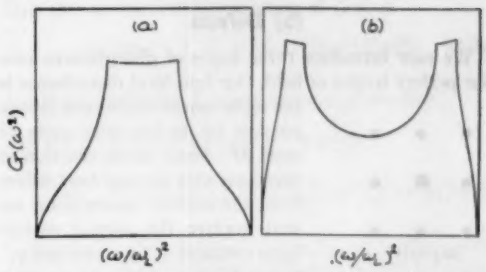


FIG. 2. The distribution function $G(\omega^2)$ for the number of frequencies between ω^2 and $\omega^2+d\omega^2$. (a) $\gamma_1=\gamma_2=\gamma_3$, (b) $\gamma_1>\gamma_2=\gamma_3$. $G(\omega^2)$ has the same shape as the energy distribution function for the tight binding model of electrons in semiconductors.

so that as $N\to\infty$

$$f(v)=\pi^{-3}\prod_{j=1}^{3}\int_0^{\pi}\exp[2ivM^{-1}\gamma_j(1-\cos\varphi_j)]d\varphi_j$$

$$=\prod_{j=1}^{3}\{\exp(2ivM^{-1}\gamma_j)\}\{J_0(2vM^{-1}\gamma_j)\}, \tag{1.9}$$

where $J_0(x)$ is the zeroth-order Bessel function.

The distribution function $G(\omega^2)$ of ω^2 is then

$$G(\omega^2)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-iv\omega^2)f(v)dv. \tag{1.10}$$

It is related to the frequency distribution function $g(\omega)$ by $2\omega G(\omega^2)=g(\omega)$. The integration of (1.10) has been discussed by Rosenstock and Newell[10] and one of the authors.[11] It is plotted in Fig. 2.

We shall, in the usual language of solid state physics, frequently refer to the continuum of possible frequencies of our lattice as the "frequency band." We shall see later that localized disturbances in the lattice sometimes give rise to new discrete frequencies which are displaced out of the band (Fig. 3).

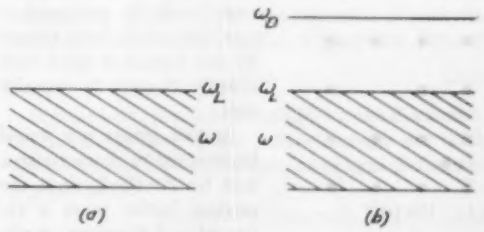


FIG. 3. Frequency spectrum of (a) perfect lattice, (b) lattice with defect. ω_L = maximum frequency, ω_D = discrete frequency for localized mode.

[10] H. B. Rosenstock and G. F. Newell, J. Chem. Phys. 21, 1607 (1953).

[11] E. W. Montroll, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, Berkeley, 1955).

(b) Defects

We now introduce three kinds of disturbances into our perfect lattice or field. Our first local disturbance is the replacement of a normal lattice particle by an impurity atom of mass M'. Since the interaction of this mass with its neighbors differs from the normal interactions, we shall replace the normal central force constant γ_1 by a new one γ_1'. Generally we shall characterize the impurity by the two parameters

impurity

$$P=\gamma_1'/\gamma_1 \quad \text{and} \quad Q=M'/M.$$

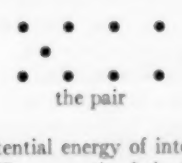
Our second type of local variation is the interstitial, a normal or impurity atom inserted between normal lattice points.

interstitial

Generally the static distortion caused by the insertion of an interstitial or an impurity extends over several lattice spacings. However, since our first aim is to get a qualitative understanding of the influence of local disturbances, we shall suppose that only nearest neighbors are directly influenced (of course more distant neighbors will be affected indirectly through new localized modes of vibration or from scattering of waves by the disturbance).

Our third type of field singularity is the hole or the omission of an atom from the lattice. It is to be noted that a hole and interstitial would generally be created together (say by radiation damage in a crystal). This might be the analog of pair production in a quantum field. Indeed in a neutral ionic lattice (or in the corresponding discussion of electron levels in semiconductors) the pair is characterized by two regions of space with average charges of opposite sign.

the hole


the pair

At first glance one would be tempted to characterize a hole by $P=Q=0$, since the original lattice mass is removed and therefore no potential energy of interaction exists with its neighbors. However, the hole forms an interesting model of a particle in a field if we characterize it by $P>1$ and $Q=0$. This choice of P is seen to resemble a hole in a crystal lattice if one examines what happens when an atom is removed from a lattice site in a monatomic lattice. The repulsive force which kept its neighbors in their places is removed so that a tendency exists for them to fall into the hole. We characterize this tendency by coupling the neighbors of the hole to the hole's lattice point by a stronger than normal spring ($\gamma_1'>\gamma_1$), or by an alternative model in which the neighbors on opposite sides of the hole are coupled together with a spring constant $\gamma_1''=\frac{1}{2}\gamma_1'>\frac{1}{2}\gamma_1$.

An interesting property of a hole is that it is the source of a pulsating mode. In a quantum mechanical system there exists a nonvanishing probability (even when the motion of the lattice is its zero-point oscillations) for a neighboring particle to fall into a hole (see Fig. 4). It will be assumed that even though it may take a long time for a neighboring particle to decide to jump into a hole, the time required for the jump to occur, when once started, is small compared to the period of the pulsating mode of the lattice. Hence the motion of a hole does not change the structure of the mode (see Fig. 4) but eventually causes a readjustment of the amplitude of the vibrations of neighboring lattice points. This readjustment is propagated through the lattice with a finite velocity. We see that the motion of a hole is a random process. In a very small time the hole either remains stationary or jumps to one of its neighboring lattice positions. The existence of an external field or other holes might give a preference to jump in a particular direction.

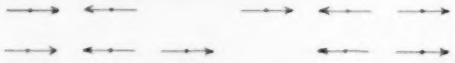


FIG. 4. The pulsating localized mode for a hole. The motion of the hole (illustrated above as a displacement of one lattice spacing to the right) does not alter the character of the mode.

Classically this random process corresponds to the random walk of a particle on a lattice. In the limit of very small lattice spacings its motion can sometimes be described through the diffusion equation. If, however, the lattice and hole are disturbed by the measurements required for the determination of transition probabilities of various jumps, a Schrödinger equation yields a more appropriate description in this limit. The effective mass of the hole would be assigned in the manner that is used to define the effective mass of an electron in a metal or semiconductor or of a bubble in a fluid. This paper is mainly concerned with one-dimensional systems. Although formulas will be given for many three-dimensional situations, the details of these more lengthy calculations will be reserved for a subsequent paper.

2. GENERAL EQUATIONS FOR A ONE-DIMENSIONAL LATTICE

The equations of motion of a one-dimensional ring of N atoms are

$$M\ddot{x}_j-\gamma(x_{j+1}-2x_j+x_{j-1})=D_jx_j, \qquad j=-\tfrac{1}{2}N+1, \cdots, \tfrac{1}{2}N, \quad (2.1a)$$

$$x_{N+j}=x_j. \quad (2.1b)$$

Here x_j represents the displacement of the jth atom from its equilibrium position and D_j is a difference differential operator which characterizes the lattice defects.

In the perfect lattice ($D_j \equiv 0$) of lattice spacing a, the displacements can be expressed as a linear combination of the normal modes

$$x_j = A_k a \exp i(k\omega_k + j\varphi_k); \quad \varphi_k = 2\pi k/N \tag{2.2a}$$

(and those with exponentials of $-k\omega_k$). Here

$$\omega_k = 2(\gamma/M)^{\frac{1}{2}} \sin\tfrac{1}{2}\varphi_k = \omega_L \sin\tfrac{1}{2}\varphi_k, \tag{2.2b}$$

where ω_L is the largest frequency,

$$M\omega_L^2 = 4\gamma. \tag{2.2c}$$

The dimensionless constants $\{A_k\}$ are to be chosen to satisfy initial conditions in the lattice.

The quantum mechanical zero-point energy of our perfect lattice is

$$E_0 = \tfrac{1}{2}\hbar\omega_L \sum_{k=1}^{N} \sin(\pi k/N) \simeq \hbar\omega_L N/\pi, \tag{2.3}$$

so that the zero-point per lattice point is $\epsilon_0 = \hbar\omega_L/\pi$.

When lattice defects exist the normal modes divide into two classes; those whose frequencies suffer slight variations in the band, and those whose frequencies are displaced out of the band. This latter group corresponds to modes localized around the defects and must be analyzed separately. The former group can be discussed by perturbation theory. The detailed manner in which this division into classes occurs is presented in Appendix I.

When perturbation theory is not appropriate we assume that

$$x_j = aAu(j) \exp -i\omega t, \tag{2.4}$$

where $u(j)$ is a set of dimensionless numbers which in an infinite lattice are defined by

$$Lu(j) = M\omega^2 u(j) + \gamma[u(j+1) - 2u(j) + u(j-1)]$$
$$= \sum_k w^{(k)}(k+j)u(k+j) \tag{2.5a}$$
$$j = 0, \pm 1, \pm 2, \cdots,$$

and which are localized around defects so that

$$u(j) \to 0 \quad \text{as} \quad j \to \pm\infty. \tag{2.5b}$$

The operator L is defined so that $Lu(j)$ gives the middle quantity in (2.5a). It has the differential representation $L = M\omega^2 - 2\gamma(1 - \cosh D)$; $D = d/dj$. The $w^{(k)}(j+k)$'s characterize the lattice defects and vanish as $j \to \pm\infty$.

The general solution of (2.5a) can be obtained in terms of the Green's function $g(j)$ which is defined by

$$Lg(j) = \begin{cases} \gamma & \text{if } j = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

It is easy to see that the solution of (2.6) is

$$u(j) = \gamma^{-1} \sum_k \sum_l g(j+k-l) w^{(k)}(l) u(l) \tag{2.7a}$$

for, if we apply the L operator to $u(j)$ we find

$$Lu(j) = \gamma^{-1} \sum_k \sum_l w^{(k)}(l) u(l) Dg(j+k-l)$$
$$= \sum_{k,l} w^{(k)}(l) u(l) \delta(j+k-l) = \sum_k w^{(k)}(j+k) u(j+k).$$

If the disturbance has an influence on lattice points $l_1, l_2, \cdots, l_s$ then the new frequencies and normal modes are found by successively setting j (in 2.7) equal to $l_1, l_2, \cdots, l_s$ and finding the values of ω for which this set of homogeneous equations has solutions. It is to be noted that the Green's function $g(j)$ depends on ω. When the solutions of these equations have been found, we see that

$$u(j) = \gamma^{-1} \sum_{\beta=1}^{s} \sum_k g(j+k-l_\beta) w^{(k)}(l_\beta) u(l_\beta). \tag{2.7b}$$

The Green's function which satisfies (2.6) is

$$g(j) = (\gamma/\pi) \int_0^\pi \frac{\cos\varphi j d\varphi}{2\gamma(\cos\varphi - 1) + M\omega^2} \tag{2.8a}$$
$$= \frac{1}{2\pi} \int_0^\pi \frac{\cos j\varphi d\varphi}{\cos\varphi + \cosh z},$$
$$= \frac{\exp\{-|j|(z+i\pi)\}}{2\sinh z} = \frac{(-1)^j \exp(-|j|z)}{2\sinh z}, \tag{2.8b}$$

where, defining f as $f = \omega/\omega_L$,

$$\cosh\tfrac{1}{2}z = \omega/\omega_L = f > 1. \tag{2.9}$$

Note that (2.5b) implies

$$g(j) \to 0 \text{ as } j \to \infty \text{ and therefore that } z > 1. \tag{2.10a}$$

Also, since $z \to \infty$ as $f \to \infty$,

$$g(j) \to 0 \text{ as } f \to \infty \text{ for fixed } j. \tag{2.10b}$$

This means that the higher the frequency of a localized mode, the more rapidly it decays spacewise. Note also

$$g(j) = g(-j). \tag{2.10c}$$

Similar formulas can be derived for the discussion of the propagation of a plane wave through our lattice. Let us assume that in the absence of local disturbances a plane wave

$$x_j = aA \exp i(j\varphi_0 - \omega t), \tag{2.11a}$$

with

$$M\omega^2 = 2\gamma(1 - \cos\varphi_0) \quad \text{or} \quad \omega = 2(\gamma/M)^{\frac{1}{2}} \sin\tfrac{1}{2}\varphi_0 \tag{2.11b}$$

is propagated through the lattice. Here Aa represents a length which is a fraction A of a lattice spacing. If the "scattering" of the plane wave by the disturbance is elastic, the complete displacement of the jth lattice

particle is

$$x_j = A\alpha e^{-i\omega t}[u(j)+\exp ij\varphi_0], \tag{2.12}$$

where $u(j)$ is the solution of the equations

$$Lu(j) = M\omega^2 u(j)+\gamma[u(j+1)-2u(j)+u(j-1)]$$
$$= \sum_k w^{(k)}(k+j)[u(k+j)+\exp i(k+j)\varphi_0], \tag{2.13}$$

the disturbance again being introduced through the set of numbers $\{w^{(k)}(k+j)\}$. The solution of this set of equations is obtained as before, through the use of the Green's function $g(j)$ which satisfies (2.6). In this case the denominator of (2.8a) has a pole (since now $0<\omega<\omega_L$). Hence we must use the Cauchy principal value of (2.8a) to find

$$g(j) = (\gamma/\pi)\oint_0^\pi \frac{\cos\varphi j d\varphi}{2\gamma(\cos\varphi-1)+M\omega^2}$$
$$= \frac{1}{2\pi}\oint_0^\pi \frac{\cos\varphi j d\varphi}{\cos\varphi-\cos\varphi_0} = -\frac{1}{2}\frac{\sin(|j|\varphi_0)}{\sin\varphi_0}. \tag{2.14}$$

The solution of (2.13) is easily seen to be

$$u(j) = \gamma^{-1}\sum_{k,l} g(j+k-l)w^{(k)}(l)[u(l)+\exp il\varphi_0]. \tag{2.15}$$

Since the postulated disturbance is chosen to be localized, $w^{(k)}(l)$ vanishes for only a small number of values of l, and the displacements in the summand of (2.15) range over only a small set of l's. If these be l_1, l_2, $\cdots$, l_s then we must determine $u(l_1)$, $\cdots$, $u(l_s)$ by succesively letting j be l_1, $\cdots$, l_s and solving the resulting s inhomogeneous equations for $u(l_1)$, $\cdots$, $u(l_s)$. Then the final solution of (2.15) is

$$u(j) = \gamma^{-1}\sum_k \sum_\beta g(j+k-l_\beta)w^{(k)}(l_\beta)[u(l_\beta)+\exp i\varphi_0 l_\beta]. \tag{2.16}$$

If we let u be a vector with components $u(l_\beta)$, $\beta=1, \cdots, s$ we see that the s equations derived from (2.16) for $u(l_1), \cdots, u(l_s)$ can be written in the matrix form

$$u = Gu+Gv, \tag{2.17}$$

where v is a known vector and G a known matrix. Then, if ψ_μ and ζ_μ are the characteristic vectors of G defined by

$$G\psi_\mu = \lambda_\mu\psi_\mu \quad \text{and} \quad \zeta_\mu G = \lambda_\mu\zeta_\mu \tag{2.18a}$$

$$\zeta_\mu\cdot\psi_\nu = \delta_{\mu\nu}, \tag{2.18b}$$

where $\{\lambda_\mu\}$ is the set of eigenvalues of G, we can expand the known vector v as

$$v = \sum v_\mu\psi_\mu; \quad v_\mu = \zeta_\mu\cdot v. \tag{2.19}$$

Then it is easily seen that

$$u = \sum_\mu [\lambda_\mu v_\mu/(1-\lambda_\mu)]\psi_\mu. \tag{2.20}$$

It is to be noted that the characteristic values and vectors of G can also be used to find the frequencies of localized modes. Equation (2.7b) is equivalent to

$$u = Gu,$$

when $j=l_1, l_2, \cdots, l_s$. Hence if we express u as $u=\sum u_\mu\psi_\mu$ the condition for the existence of solutions of $u=Gu$ is

$$\sum u_\mu(1-\lambda_\mu)\psi_\mu = 0 \quad \text{or} \quad u_\mu(1-\lambda_\mu)=0 \text{ for all } \mu. \tag{2.21}$$

We shall now apply the various formulas derived in this section to the investigation of the effect of an isolated defect in a one-dimensional lattice.

3. ISOLATED POINT DEFECT

This section is concerned with the theory of localized modes associated with isolated defects. The amplitude of lattice displacements due to these modes diminishes exponentially with distance from the defect. The types of defects analyzed are

(a) The variation of the mass of a single particle and of the force constants associated with that particle and its nearest neighbors. If the force constants are not changed this corresponds to the introduction of an isotope into the lattice.
(b) The interstitial.
(c) The changing of a single force constant.

The topology of a linear chain is such that the localized modes of an interstitial can be deduced from those of a normal impurity, and that of a hole can be deduced from the results of (c).

In general the localized modes associated with impurities are either symmetrical or antisymmetrical about the position of the impurity (particles to the left being either in or 180° out of phase with those to the right). As the mass of the impurity is reduced the frequency of the symmetric mode increases and its range of influence on its neighbors decreases; in the limit of zero mass the effect of this mode is no longer propagated through the lattice. The only mode associated with a hole is the antisymmetric one; this mode exists independently of the mass of the impurity provided the new spring constant is sufficiently large (in fact if $\gamma'/\gamma = P > 2$). The impurity frequencies are plotted in Fig. 7.

A brief discussion of the scattering by an isotope is given at the end of the section.

Let us insert an impurity atom of mass M' in the lattice point "0" in a chain of $2N+1$ atoms, and let γ' be the force constant of the "spring" which connects this atom to atoms "1" and "−1." Then Eqs. (2.5) become

$$Lu(j) = 0 \text{ except when } j=0, \pm 1, \tag{3.1a}$$

$$Lu(0) = (M-M')\omega^2 u(0) + (\gamma-\gamma')[u(1)-2u(0)+u(-1)], \tag{3.1b}$$

$$Lu(-1) = (\gamma'-\gamma)[u(-1)-u(0)], \tag{3.1c}$$

$$Lu(1) = (\gamma'-\gamma)[u(1)-u(0)], \tag{3.1d}$$

and the w's of (2.5) are

$$w^{(k)}(k+j)=0, \quad \text{for all } k \text{ and } j \text{ except} \tag{3.2a}$$

$$w^{(0)}(0)=(M-M')\omega^2-2(\gamma-\gamma'); \qquad w^{(\pm 1)}(\pm 1)=\gamma-\gamma', \tag{3.2b}$$

$$w^{(1)}(0)=w^{(-1)}(0)=\gamma-\gamma'; \quad w^{(0)}(\pm 1)=\gamma'-\gamma. \tag{3.2c}$$

The solution of Eqs. (2.5) from which localized normal modes are to be determined is

$$\begin{aligned} u(j)=\gamma^{-1}u(0)\{&g(j)[(M-M')\omega^2-2(\gamma-\gamma')] \\ &+(\gamma-\gamma')[g(j+1)+g(j-1)]\} \\ &+u(1)(1-P)[g(j)-g(j-1)] \\ &+u(-1)(1-P)[g(j)-g(j+1)], \end{aligned} \tag{3.3}$$

where we set

$$P=\gamma'/\gamma \quad \text{and} \quad Q=M'/M, \quad f=\omega/\omega_L. \tag{3.4}$$

Since the coefficient of $u(0)$ will appear in many places we employ the definition of $g(j)$ [Eq. (2.6)] to derive it in a more compact form:

$$\begin{aligned} &\gamma^{-1}\{g(j)[(M-M')\omega^2-2(\gamma-\gamma')] \\ &\qquad +(\gamma-\gamma')[g(j+1)+g(j-1)]\} \\ &\qquad =\begin{cases} 4f^2g(j)(P-Q) & \text{if } j\neq 0 \\ 4f^2g(0)(P-Q)+(1-P) & \text{if } j=0. \end{cases} \end{aligned} \tag{3.5}$$

The matrix G defined in (2.17) is

$$G=\begin{bmatrix} (1-P)[g(1)-g(0)] & 4f^2g(1)(P-Q) & (1-P)[g(1)-g(2)] \\ (1-P)[g(0)-g(1)] & 4f^2g(0)(P-Q)+(1-P) & (1-P)[g(0)-g(1)] \\ (1-P)[g(1)-g(2)] & 4f^2g(1)(P-Q) & (1-P)[g(1)-g(0)] \end{bmatrix}. \tag{3.6}$$

One of the characteristic vectors of G is

$$\psi_1=\{1, 0, -1\} \text{ with } \lambda_1=(1-P)[g(2)-g(0)]. \tag{3.7a}$$

The adjoint (left-hand characteristic vector of G) is

$$\zeta_1=\tfrac{1}{2}(1, 0, -1), \text{ so that } \zeta_1\cdot\psi_1=1. \tag{3.7b}$$

This is a pulsating mode

$$\leftarrow \cdot \rightarrow. \tag{i}$$

It is the one-dimensional analog of a periodically expanding and contracting spherical bubble.

The other solutions are easily found when $P=1$ (the case of an impurity such as an isotope which does not distort the lattice, $\gamma'=\gamma$). Then

$$\psi_2=\{1, g(0)/g(1), 1\}; \quad \lambda_2=4f^2(1-Q)g(0), \tag{3.8}$$

and

$$\zeta_2=(0, g(1)/g(0), 0). \tag{3.9}$$

Finally

$$\psi_3=\{1,0,1\}, \quad \lambda_3=0, \tag{3.10}$$

$$\zeta_3=\tfrac{1}{2}(1, 2g(1)/g(0), 1). \tag{3.11}$$

The modes in this case are

$$\rightarrow \leftarrow \rightarrow, \tag{ii}$$

$$\rightarrow \cdot \rightarrow. \tag{iii}$$

In the general case the roots λ_2 and λ_3 satisfy the characteristic equation

$$\begin{aligned} \lambda^2-\lambda(&(1-P)\{[g(1)-g(2)]+3[g(1)-g(0)]\} \\ &+(1-Q)\{1-2[g(1)-g(0)]\}) \\ &+4f^2g(1)(1-P)(1-Q)=0. \end{aligned} \tag{3.12}$$

The characteristic vectors are of the form

$$\psi=\{a,b,a\}, \tag{3.13a}$$

with an a/b ratio

$$\frac{a}{b}=\frac{\lambda+P-1-4f^2g(0)(P-Q)}{2(1-P)[g(0)-g(1)]}. \tag{3.13b}$$

The left characteristic vectors are of the form

$$\zeta=(2ac+bd)^{-1}(c,d,c), \tag{3.14a}$$

with

$$\frac{c}{d}=\frac{\lambda+P-1-4f^2g(0)(P-Q)}{8f^2g(1)(P-Q)}. \tag{3.14b}$$

Since there are two roots of our characteristic equation these are the vectors of the form (3.13) and (3.14).

The frequencies associated with our isolated defect modes are determined by referring back to (2.21). In order that

$$u=\{u(-1), u(0), u(1)\} \tag{3.15a}$$

be expressible as

$$u=\sum u_\mu\psi_\mu, \tag{3.15b}$$

the condition $(1-\lambda_\mu)u_\mu=0$ must be satisfied for all μ. Hence if $u\neq 0$ then $\lambda_\mu=1$. The frequency ω associated with the symmetrical pulsating mode (i) [Eq. (3.7a)] is then determined by

$$(1-P)[g(2)-g(0)]=1. \tag{3.16}$$

When $\omega/\omega_L>1$, $g(j)$ is given by (2.8b). Therefore we have

$$\tfrac{1}{2}(1-P)(e^{-2z}-1)/\sinh z=1 \quad \text{with } z>1 \tag{3.17}$$

so that $\exp(-z)=1/(P-1)$. This quantity is <1 as is required only when $P=\gamma'/\gamma>1$. Then

$$(\omega/\omega_L)^2=\cosh^2\tfrac{1}{2}z=-\frac{1}{4}\frac{P^2}{(P-1)} \quad \text{when } P>2; \tag{3.18}$$

as $P\to 2$ this mode returns to the band.

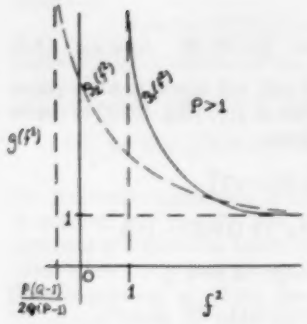


FIG. 5. The intersection of $g_1(f^2)$ and $g_2(f^2)$ represents the solution of Eq. (3.22). $P>1$.

The displacement of the jth lattice point of a system oscillating in this mode is [see Eq. (2.76)]

$$u(j)=(1-P)u(1)[1/(1-P)]^j, \quad j\geq 1 \tag{3.19a}$$

$$u(j)=-u(-j) \quad \text{for } j<0. \tag{3.19b}$$

Note that large values of P correspond to high frequencies and a rapid drop in $|u(j)|$ as $|j|$ increases. Also, the frequency is independent of the mass of the defect particle.

The symmetric mode (ii) $[u(j)=u(-j)]$ has an easily determined frequency when $P=1$ (the isotope case). Then (3.8) implies

$$4(\omega/\omega_L)^2=1/g(0)(1-Q),$$

or

$$(\omega/\omega_L)^2=1/Q(2-Q) \quad \text{with } 0<Q<1. \tag{3.20}$$

This mode goes back into the band as $Q\to 1$. It has the property

$$u(j)=u(-j)=[Q/(Q-2)]^{|j|}u(0) \quad \text{when } 0<Q<1. \tag{3.21}$$

It is interesting to note that the frequency (3.20) is the same as that of the pulsating mode of a hole if one replaces P by $2/Q$.

The frequency of the symmetric mode (ii) is obtained as a general function of P and Q by setting $\lambda=1$ in Eq. (3.12). If one uses (2.8b), (2.9), and the equivalent of (2.9), $\sinh\frac{1}{2}z=(f^2-1)^{\frac{1}{2}}$ with $f=\omega/\omega_L$ some algebraic manipulation yields the following equation for the frequency of our required symmetrical mode:

$$\left(\frac{f^2}{f^2-1}\right)^{\frac{1}{2}}=1+\frac{QP}{P(1-Q)+2Q(P-1)f^2}, \tag{3.22}$$

while the factor (b/a) in (3.13) becomes

$$a/b=-Qf(f-[f^2-1]^{\frac{1}{2}}). \tag{3.23}$$

An asymptotic expression for the frequency $f=\omega/\omega_L$ can easily be obtained in the limit of very small mass of the impurity $Q=\epsilon\simeq 0$. Here we assume $f^2=a/Q$ and find a by noting that

$$[f^2/(f^2-1)]^{\frac{1}{2}}\simeq 1+Q/2a+O(Q^2), \tag{3.24}$$

so that (3.22) becomes

$$Q/2a=O(Q^2)=QP/[P+2a(P-1)]+O(Q^2),$$

or $a=\frac{1}{2}P$ and as $Q\to 0$

$$f^2\sim\tfrac{1}{2}P/Q+O(Q). \tag{3.25}$$

Hence as the relative mass $Q=M'/M$ of an impurity vanishes, the frequency of the local mode becomes very large while the effective range vanishes. In the limit $a=0$ the neighboring lattice points do not feel the effect of this mode and the particle of mass M' disappears leaving a hole. One degree of freedom of our chain then disappears. Then the only localized mode associated with a hole is the antisymmetric or pulsating mode (i). It can be shown that a/b in Eq. (3.23) satisfies $-1<a/b<0$ when $f>1$ and $1>Q>0$.

Our frequency ratio f is displaced out of the band only when certain relations exist between P and Q. One way of solving (3.22) is to plot both the left and right hand sides as functions of f^2 and to find their point of intersection (see Fig. 5). We shall denote $[f^2/(f^2-1)]^{\frac{1}{2}}$ by $g_1(f^2)$ and the right hand side by $g_2(f^2)$. Two cases are to be examined separately, namely $P>1$ and $0<P<1$ ($P=1$ having already been disposed of). It will generally be assumed that $Q\geq 0$. When $P>1$ the right-hand side of (3.22) approaches $1+P[2f^2(P-1)]^{-1}>1+\frac{1}{2}f^{-2}$. Hence, the function $g_2(f^2)$ eventually lies above $g_1(f^2)$ as $f^2\to\infty$. Since $g_2(f^2)$ forms a hyperbola with one asymptote at $f^2=P(Q-1)/2Q(P-1)$ and the other at $f^2=1$, the condition for the interaction of $g_1(f^2)$ and $g_2(f^2)$ to exist (as it must if our required frequency is to rise out of the band) is that

$$P(Q-1)/2Q(P-1)<1 \quad \text{or} \quad P\geq Q(2-P).$$

Otherwise the asymptote of $g_2(f^2)$ lies to the right of that of $g_1(f^2)$ and no intersection or proper value of (3.22) exists.

When $P<1$ the hyperbola which represents $g_2(f^2)$ is twisted by 90° about the intersection of the asymptotes (see Fig. 6). Here an intersection of g_1 and g_2 exists as long as

$$(Q-1)P/2(P-1)Q>1 \quad \text{or} \quad P>Q(2-P)$$

as in the foregoing. It is to be recalled from (3.20) that a solution exists when $P=1$ if $0<Q<1$. Hence, our

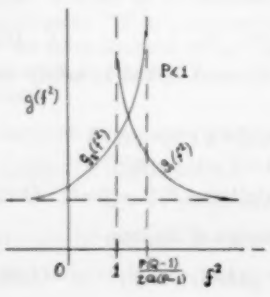


FIG. 6. The intersection of $g_1(f^2)$ and $g_2(f^2)$ represents the solution of Eq. (3.22). $P<1$.

result can be summarized by stating that a normal mode of type (ii) exists with a frequency above the band when all the following conditions are satisfied

$$P>0, \quad P>Q(2-P), \quad \text{and} \quad Q\geqslant 0 \qquad (3.26)$$

with the frequency going to infinity as $Q\to 0$.

Equation (3.22) can be squared and rearranged to yield the following quadratic equations for the squares of the normal mode frequencies

$$4Q^2(P-1)f^4+PQf^2(4-PQ-2P)-P^2=0. \qquad (3.27)$$

The reader is to be warned that spurious roots generally exist. The only root of significance here has the property $(\omega/\omega_L)^2\geqslant 1$. A root can always be checked for appropriateness by substituting the corresponding values of f^2 back into (3.22). The impurity frequencies are plotted in Fig. 7.

Once the frequencies of modes of type (ii) are known, all displacements can be discussed in terms of $u(0)$ through the equation

$$u(j)=Qu(0)(1+e^z)\exp-|j|(z+i\pi);$$
$$\omega/\omega_L=\cosh\tfrac{1}{2}z. \qquad (3.28)$$

As $(\omega/\omega_L)\to\infty$, $z\to\infty$ and $u(j)\to 0$ so that high frequencies imply rapid spacewise decay as was mentioned earlier.

The characteristic equation (3.12) may have two roots with $\lambda=1$ when special relations exist between P and Q. However, this is not the case under two conditions of greatest interest; those with either $P=1$ (the "isotope" case) or $Q=0$ (the case of the "hole").

A discussion of the interstitial proceeds along the same line that was followed above for holes and impurities. Since an interstitial does not change the topology of a linear chain one expects the results to be equivalent to those derived in the foregoing (as can indeed be proven to be the case). Since an interstitial in a three dimensional simple cubic lattice gives a local body centered structure, this equivalence disappears in a real lattice.

We shall now give a brief analysis of the interstitial to show how it can be fitted into the mathematical formalism discussed in Sec. 2, even though no unexpected results are obtained.

Consider a linear chain of $2N$ atoms and suppose an

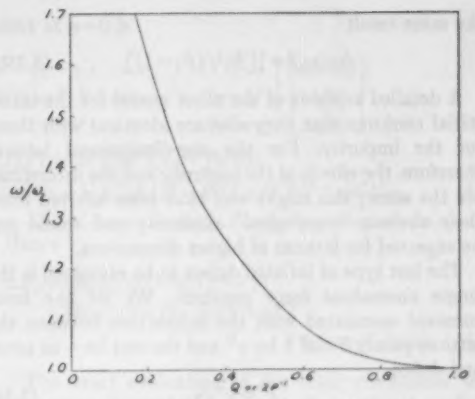


FIG. 7. Variation of impurity frequency ω with impurity parameter ($Q=M'/M$ for an isotope, $P=\gamma'/\gamma$ for a hole). ω_L = maximum frequency for the perfect lattice.

atom of mass M'' is inserted at an interstitial site "½" between the two sites "0" and "1." If γ'' is the force constant of the spring connecting the interstitial to its neighbors, then the fundamental equations (2.5) become

$$Lu(j)=0 \quad \text{except when } j=0,1 \qquad (3.29a)$$
$$Lu(j)=(\gamma''-\gamma)u(0)-\gamma''u(\tfrac{1}{2})+\gamma u(1) \quad \text{if } j=0, \qquad (3.29b)$$
$$Lu(j)=(\gamma''-\gamma)u(1)-\gamma''u(\tfrac{1}{2})+\gamma u(0) \quad \text{if } j=1, \qquad (3.29c)$$

with the additional equation

$$\gamma u(\tfrac{1}{2})=-\gamma''[u(1)-2u(\tfrac{1}{2})+u(0)]$$
$$+(\gamma-M''\omega^2)u(\tfrac{1}{2}). \qquad (3.29d)$$

The form of this last equation is somewhat arbitrary; the present choice is consistent with the definition (2.6) of the Green's function. The w functions are all zero except that, with an obvious extension of the notation

$$w^{(0)}(0)=w^{(0)}(1)=\gamma''-\gamma, \qquad (3.30a)$$
$$w^{(1)}(1)=w^{(-1)}(0)=\gamma, \qquad (3.30b)$$
$$w^{(\frac{1}{2})}(\tfrac{1}{2})=w^{(\frac{1}{2})}(1)=w^{(-\frac{1}{2})}(\tfrac{1}{2})=w^{(-\frac{1}{2})}(0)=-\gamma'', \qquad (3.30c)$$
$$w^{(0)}(\tfrac{1}{2})=2\gamma''+(\gamma-M''\omega^2). \qquad (3.30d)$$

The determination of the normal modes and frequencies depends on the solution of the eigenvalue problem for the matrix G, given in the present case by

$$G=\begin{bmatrix}(P_I-1)g(0)+g(1) & -P_I[g(0)+g(1)] & g(0)+(P_I-1)g(1)\\ \tfrac{1}{2}P_I & 1-P_I+2Q_If^2 & \tfrac{1}{2}P_I\\ (P_I-1)g(1)+g(0) & -P_I[g(0)+g(1)] & g(1)+(P_I-1)g(0)\end{bmatrix}, \qquad (3.31a)$$

where

$$P_I=\gamma''/\gamma \quad \text{and} \quad Q_I=M''/M. \qquad (3.31b)$$

Although this matrix seems quite different from that given by (3.6) for the impurity problem, it leads to the same normal modes and frequencies. For example, one characteristic vector of G is

$$\psi_1=\{1,0,-1\} \text{ with } \lambda_1=(P_I-2)[g(0)-g(1)]. \qquad (3.32)$$

The corresponding normal frequency ω is obtained from

$$(P_I-2)[g(0)-g(1)]=1, \qquad (3.33a)$$

which is only an alternative form of (3.16) leading to

the same result

$$(\omega/\omega_L)^2=\tfrac{1}{4}[P_I^2/(P_I-1)]. \tag{3.33b}$$

A detailed analysis of the other modes for the interstitial confirms that they also are identical with those for the impurity. For the one-dimensional lattice, therefore, the effects of the impurity and the interstitial are the same; this might well have been inferred from their obvious "topological" similarity and would not be expected for lattices of higher dimensions.

The last type of isolated defect to be examined is the single anomalous force constant. We let the force constant associated with the interaction between the lattices points 0 and 1 be γ'' and the rest be γ as usual and define

$$P'=\gamma''/\gamma. \tag{3.34}$$

The required equations of motion are

$$Lu(j)=0 \quad \text{except when } j=0,\ 1, \tag{3.35a}$$

$$Lu(j)=(\gamma-\gamma'')[u(1)-u(0)] \quad \text{for } j=0, \tag{3.35b}$$

$$Lu(j)=(\gamma-\gamma'')[u(0)-u(1)] \quad \text{for } j=1. \tag{3.35c}$$

Hence,

$$w^{(k)}(k+j)=0 \quad \text{for all } k \text{ and } j \text{ except}$$
$$w^{(1)}(1)=w^{(-1)}(0)=(\gamma-\gamma'')=-w^{(0)}(0)=-w^{(0)}(1). \tag{3.36}$$

The G matrix is

$$G=(1-P')[g(1)-g(0)]\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}. \tag{3.37}$$

The characteristic values of G are

$$\lambda_1=0, \quad \lambda_2=2(1-P')[g(1)-g(0)], \tag{3.38}$$

with characteristic vectors

$$\zeta_1=\tfrac{1}{2}(1,\ 1), \qquad \psi_1=\{1,\ 1\}, \tag{3.39a}$$

$$\zeta_2=\tfrac{1}{2}(1,\ -1), \quad \psi_2=\{1,\ -1\}. \tag{3.39b}$$

The antisymmetrical vector ψ_2 yields a frequency out of the band, for if we let $\lambda_2=1$ and define $g(j)$ by (2.8b) we have

$$f^2=(\omega/\omega_L)^2=(P')^2/(2P'-1), \tag{3.40}$$

which is exactly the same as the frequency of the pulsating mode of a hole [Eq. (3.18)] if we identify P' with $\frac{1}{2}P$. The decay factor $\exp-\varepsilon$ is

$$\exp-\varepsilon=1/(2P'-1). \tag{3.41}$$

The theory of scattering of plane waves which is summarized in Eqs. (2.16)–(2.00) can be applied to the problem of scattering by an isolated defect. The simplest case is that of the isotope, $P=1$. Equation (2.16) reduces to

$$u(j)=\gamma^{-1}g(j)\omega^2(M-M')[u(0)+1], \tag{3.42}$$

where $g(j)$ is (see 2.14)

$$g(j)=\tfrac{1}{2}(\sin|j|\varphi_0)/\sin\varphi_0, \tag{3.43}$$

with

$$\omega/\omega_L=f=\sin\tfrac{1}{2}\varphi_0, \tag{3.44}$$

and $u(0)$ is found by setting $j=1$. Since $g(0)=0$, $u(0)=0$ and

$$u(j)=2f^2(1-Q)\sin(|j|\varphi_0)/\sin\varphi_0. \tag{3.45}$$

The total displacement of the jth lattice point from its equilibrium position is (see Eq. 2.12)

$$x_j=aAe^{-i\omega t}[e^{ij\varphi_0}+2f^2(1-Q)\sin(|j|\varphi_0)/\sin\varphi_0]. \tag{3.46}$$

Since the term $\sin|j|\varphi_0$ contains terms with both positive and negative exponents, the scattered wave contains both a reflected and transmitted part.

The full matrix theory must be used when $P\neq 1$.

4. ZERO-POINT OR VIBRATIONAL SELF-ENERGY OF DEFECTS

An important characteristic of a lattice defect is its vibrational self-energy. We define this quantity as the difference between the zero-point energy of the lattice with a defect and that of a perfect lattice. The total potential energy of a lattice can be approximated as a constant term (the sum of the bond energies) plus a term quadratic in relative displacements. Hence the total self-energy is the sum of that required to make the change in the bond energies associated with the introduction of the defect plus the vibrational self-energy.

The vibrational self-energy can be decomposed into two parts; the first being due to the shifting of frequencies in the band, ΔE_b, and the second being due to the removal of frequencies from the band, ΔE_D. If $\{\omega_j^{(0)}\}$ represents the vibrational frequencies of a perfect lattice and $\{\omega_j^{(1)}\}$ those of a lattice with a single defect, the vibrational self-energy is

$$\Delta E=\tfrac{1}{2}\sum_j \hbar[\omega_j^{(1)}-\omega_j^{(0)}]=\Delta E_b+\Delta E_D. \tag{4.1}$$

When Q is close to 1, ΔE follows immediately from perturbation theory as well as from an exact analysis of all the normal modes (those inside as well as outside the band).

In order to avoid the necessity of using perturbation theory of degenerate characteristic values, we shall use the solutions of

$$Lu(n)=0 \quad \text{with} \quad u(N)=u(-N)=0 \tag{4.2a}$$

as the unperturbed characteristic vectors. The fixed particles at the end of the chain lead to normalized solutions

$$u(n)=N^{-1}\sin(N-n)\varphi$$

with

$$\varphi=\pi j/2N, \quad j=0,\ 1,\ 2,\ \cdots,\ 2N, \tag{4.2b}$$

with normal mode frequencies

$$M\omega^2=2\gamma(1-\cos\pi j/2N)=4\gamma\sin^2\pi j/4N. \tag{4.2c}$$

The first-order perturbation in $M\omega^2$ for an isolated defect at $n=0$ is

$$M\Delta\omega_k^2=\{u(-1),u(0),u(1)\}\begin{bmatrix}(\gamma'-\gamma) & (\gamma-\gamma') & 0\\ (\gamma-\gamma') & (M-M')[\omega_k^{(0)}]^2-2(\gamma-\gamma') & (\gamma-\gamma')\\ 0 & (\gamma-\gamma') & (\gamma'-\gamma)\end{bmatrix}\begin{bmatrix}u(-1)\\ u(0)\\ u(1)\end{bmatrix}$$
$$=(\gamma'-\gamma)\{[u(0)-u(-1)]^2+[u(0)-u(1)]^2\}+(M-M')[\omega_k^{(0)}]^2u^2(0)$$
$$=4(\gamma'-\gamma)N^{-1}\sin^2\tfrac{1}{2}\varphi_k[\cos^2(N-\tfrac{1}{2})\varphi_k+\cos^2(N+\tfrac{1}{2})\varphi_k]+(M-M')[\omega_k^{(0)}]^2N^{-1}\sin^2N\varphi_k, \quad (4.3)$$

where $\omega_k^{(0)}$ is the unperturbed frequency given by (4.2c).

The vibrational self-energy is (each frequency contributing $\frac{1}{2}\hbar\omega$ to the zero-point energy of the lattice)

$$\Delta E=\tfrac{1}{2}\hbar\sum\Delta\omega_k=\tfrac{1}{2}\hbar\sum[\omega_k-\omega_k^{(0)}]$$
$$=\tfrac{1}{2}\hbar\sum\omega_k^{(0)}\{\Delta\omega_k^2/[\omega_k^{(0)}]^2\}. \quad (4.4)$$

We see that (4.3) breaks into two subsets

$$\Delta\omega_k^2=\begin{cases}2(P-1)\omega_L^2N^{-1}\sin^2\dfrac{k\pi}{4N}\cos^2\dfrac{k\pi}{4N} & k \text{ even}\\ 2(P-1)\omega_L^2N^{-1}\sin^2\dfrac{k\pi}{4N}\sin^2\dfrac{k\pi}{4N} & \\ \qquad+[\omega_k^{(0)}]^2N^{-1}(1-Q) & k \text{ odd.}\end{cases}$$

Then

$$\Delta E=\tfrac{1}{4}\hbar N^{-1}\left\{2(P-1)\left[\sum_{k\ \text{even}}\omega_k^{(0)}\cos^2\frac{k\pi}{4N}+\sum_{k\ \text{odd}}\omega_k^{(0)}\sin^2\frac{k\pi}{4N}\right]+(1-Q)\sum_{k\ \text{odd}}\omega_k^{(0)}\right\}.$$

In the limit as $N\to\infty$, $\omega_k^{(0)}$ for a given odd k is practically equal to its value for the even number $k+1$. Hence, as $N\to\infty$,

$$\Delta E=\tfrac{1}{4}\hbar N^{-1}[2(P-1)+(1-Q)]\sum_{k=0}^{2N}\omega_k^{(0)}$$
$$=\tfrac{1}{2}\hbar\omega_L\pi^{-1}[2(P-1)+(1-Q)]. \quad (4.5)$$

Second-order perturbation theory gives terms of order $(P-1)^2$ and $(Q-1)^2$ to ΔE so that (4.5) is merely the beginning of a series in $(P-1)$ and $(Q-1)$.

The contribution of the localized defect modes to the vibrational self-energy of a defect are

$$\Delta E_D=\tfrac{1}{2}\hbar\sum_j(\omega_j-\omega_L)=\tfrac{1}{2}\hbar\omega_L\sum_j(f_j-1), \quad (4.6)$$

where the summation extends over all defect modes which go out of the band. The assumption is made that the defect mode is the displacement of a mode which was very close to the top of the band in the perfect lattice. We shall give an explicit formula only in the case of the "isotope." When $P=1$ (the "isotope" case $\gamma=\gamma'$) the only mode which escapes from the band is the symmetric one with frequency

$$f=\omega/\omega_L=[1/Q(2-Q)]^{\frac{1}{2}}.$$

Hence

$$\Delta E_D=\tfrac{1}{2}\hbar\omega_L([Q(2-Q)]^{-\frac{1}{2}}-1)$$
$$=\tfrac{1}{2}\hbar\omega_L[\{1-(1-Q)^2\}^{-\frac{1}{2}}-1]$$
("isotope" case). (4.7)

The exact evaluation of the total vibrational self-energy can be achieved through the application of the following theorem[12] on contour integration: Let $g(z)$ be a function analytic inside a contour C and $f(z)$ be a function with zeros at $z_1, z_2, \cdots, z_n$ inside C. Then, if $f(z)$ has no poles inside or on the boundary of C

$$\frac{1}{2\pi i}\int_C\frac{f'(z)}{f(z)}g(z)dz=\sum_j g(z_j). \quad (4.8)$$

Those zeros of $f(z)$ which lie on C are given a weight $\frac{1}{2}$ in the sum.

It is shown in the Appendix that the possible normal mode frequencies of a perfect chain are

$$\omega=\omega_L\sin\tfrac{1}{2}\varphi, \quad (4.9)$$

where the φ's are zeros of the two functions

$$f_1^{(0)}(\varphi)=\sin N\varphi \quad \text{and} \quad f_2^{(0)}(\varphi)=\cos N\varphi \quad (4.10)$$

in the range $0<\mathrm{Re}\,\varphi\leqslant\pi$. A lattice with a single isotopic defect of mass $M'=QM$ also has frequencies given by (4.9) where the φ's are the zeros of

$$f_1(\varphi)=\sin N\varphi, \quad (4.11a)$$

$$f_2(\varphi)=\cos N\varphi+(1-Q)\tan\tfrac{1}{2}\varphi\sin N\varphi. \quad (4.11b)$$

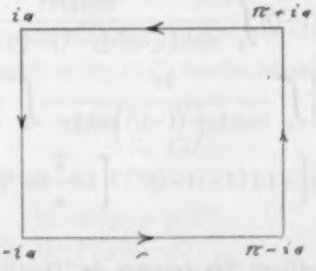


FIG. 8. The contour C which is the path of integration used in (4.12).

[12] See E. C. Titchmarsh, *Theory of Functions* (Oxford University Press, Oxford, 1932), p. 116.

The application of (4.8) implies that

$$\Delta E=\frac{\hbar\omega_L}{4\pi i}\int_C \sin\tfrac{1}{2}z\frac{d}{dz}[\log f_2(z)-\log f_2^{(0)}(z)]dz, \quad (4.12)$$

where the contour C (see Fig. 8) contains the zeros of both $f_2(z)$ and $f_2^{(0)}(z)$. As is discussed in the Appendix, the zeros of $f_2^{(0)}(z)$ and $f_2(z)$ which correspond to normal modes in the band are located on the real axis in the interval $(0,\pi)$. The frequency of the localized mode corresponds (as $N\to\infty$) to the points on the line $z=\pi+iy$ at $y=\pm\log[(2-Q)/Q]$. We take our contour, C, through both of these points since each (being on the boundary) contributes only one half of the required zero-point energy to (4.12).

It is easy to show that

$$\Delta E=\frac{\hbar\omega_L(Q-1)}{4\pi i}\int_C\left\{\frac{N\sin\tfrac{1}{2}z\cos\tfrac{1}{2}z+\tfrac{1}{2}\cos Nz\sin Nz}{\cos Nz\cos\tfrac{1}{2}z-(Q-1)\sin\tfrac{1}{2}z\sin Nz}\right\}\times\sin\tfrac{1}{2}z dz/\cos\tfrac{1}{2}z\cos Nz. \quad (4.13)$$

This integral can be decomposed into the sum of four integrals, one for each side of the rectangle in Fig. 8. The integrand of (4.13) is an odd function of y along the line $z=\pi+iy$ from $(-a,\pi)$ to (a,π). In the limit of very long chains, as $N\to\infty$, the two integrals along the longitudinal parts of the contour combine to give a contribution

$$[\hbar\omega_L(Q-1)/8\pi]\left\{\int_0^\pi\frac{\sin\tfrac{1}{2}(x-ia)dx}{\cos\tfrac{1}{2}(x-ia)[\cos\tfrac{1}{2}(x-ia)+i(Q-1)\sin\tfrac{1}{2}(x-ia)]}+\int_0^\pi\frac{\sin\tfrac{1}{2}(x+ia)dx}{\cos\tfrac{1}{2}(x+ia)[\cos\tfrac{1}{2}(x+ia)-i(Q-1)\sin\tfrac{1}{2}(x+ia)]}\right\}.$$

These integrals can be transformed easily into elementary real integrals and are found to be (in the case $\sinh\tfrac{1}{2}a+[Q-1]\cosh\tfrac{1}{2}a>0$)

$$\frac{1}{2\pi}\hbar\omega_L\{-\tan^{-1}(1/\sinh\tfrac{1}{2}a)+[1-(1-Q)^2]^{-\frac{1}{2}}\tan^{-1}([1-(1-Q^2)]^{\frac{1}{2}}/\times[\sinh\tfrac{1}{2}a+(Q-1)\cosh\tfrac{1}{2}a])\}. \quad (4.14)$$

Hence, for any fixed value of Q, the contribution of this term to (4.13) vanishes as $a\to\infty$.

The only remaining contribution to (4.13) in the limit $a\to\infty$ comes from the imaginary axis. As $N\to\infty$ the term $\frac{1}{2}\cosh Ny\sinh Ny$ in (4.13) dominates $N\sinh\frac{1}{2}y\times\cosh\frac{1}{2}y$ in this line contribution to (4.13). Hence, after applying the appropriate large N asymptotic formulas for $\cosh Ny$ and $\sinh Ny$, and letting $a\to\infty$ we reduce (4.13) to

$$\Delta E=\frac{\hbar\omega_L}{4\pi}(1-Q)\int_0^\infty\frac{\sinh\tfrac{1}{2}ydy}{\cosh\tfrac{1}{2}y[\cosh\tfrac{1}{2}y-(1-Q)\sinh\tfrac{1}{2}y]}=\frac{\hbar\omega_L}{4\pi}\left\{\int_0^\infty\frac{dy}{\cosh\tfrac{1}{2}y-(1-Q)\sinh\tfrac{1}{2}y}-\int_0^\infty\frac{dy}{\cosh\tfrac{1}{2}y}\right\}$$

$$=\tfrac{1}{4}\hbar\omega_L\left\{-1+[1-(1-Q)^2]^{-\frac{1}{2}}\left[1+\frac{2}{\pi}\sin^{-1}(1-Q)\right]\right\}. \quad (4.15)$$

As $(1-Q)\to0$ this becomes $\frac{1}{2}\pi^{-1}(1-Q)$ which is equivalent to the first-order perturbation result (4.5). This is due entirely to the shift of levels in the band. As $Q\to0$, $\Delta E\to\frac{1}{2}\hbar\omega_L(2Q)^{-\frac{1}{2}}$ which is the limit of (4.7a). To summarize, a very "weak" isotopic defect is described by the intra-band shift of frequencies while a "strong" isotopic defect, a very low mass particle introduced into the lattice, exerts its influence through the localized modes. The complete variation of the self-energy of an isotopic defect is plotted in Fig. 9 as a function of the mass ratio $Q=M'/M$.

Since an interstitial has the same topology as a point defect at a normal lattice position the contribution to its self-energy made by those frequencies which are displaced in the band can be inferred from the equations derived above.

The first-order perturbation theory of the anomalous force constant proceeds as follows:

$$M\Delta\omega_k^2=\{u(0),u(1)\}\begin{pmatrix}\gamma''-\gamma & \gamma-\gamma''\\ \gamma-\gamma'' & \gamma''-\gamma\end{pmatrix}\begin{pmatrix}u(0)\\ u(1)\end{pmatrix}$$

$$=[u(0)-u(1)]^2(\gamma''-\gamma)=4(\gamma''-\gamma)N^{-1}\sin^2\tfrac{1}{2}\varphi_k\cos^2(N+\tfrac{1}{2})\varphi_k.$$

Hence,

$$\Delta\omega_k^2=4[\omega_k^{(0)}]^2N^{-1}(P'-1)\begin{cases}\cos^2\pi k/4N & \text{if } k \text{ is even}\\ \sin^2\pi k/4N & \text{if } k \text{ is odd}\end{cases}$$

so that by following the reasoning used in the derivation of (4.5) we find

$$\Delta E=\tfrac{1}{2}\hbar(P'-1)\omega_L/\pi+O[(P'-1)^2]. \quad (4.16)$$

The pulsating mode which goes out of the band when $P'>1$ gives the contribution to the self-energy [see (4.6) and (3.40)]:

$$\Delta E_D=\tfrac{1}{2}\hbar\omega_L\{P'(2P'-1)^{-\frac{1}{2}}-1\}. \quad P'\geq1 \quad (4.17)$$

We conclude this section with a determination of the vibrational self-energy of a hole. The contribution from

the localized mode is

$$\Delta E_D=\tfrac{1}{2}\hbar\omega_L\left\{\frac{P}{2(P-1)^{\frac{1}{2}}}-1\right\} \quad \text{if } P>2. \qquad (4.18)$$

The first-order contribution ΔE_b of those frequencies which suffer small displacements in the band can be deduced from various formulas derived above. We take the following steps.

(a) Insert an anomalous bond with force constant P' into the lattice at the expense of $\Delta E_b^{(1)}\simeq\frac{1}{2}\hbar(P'-1)\omega_L/\pi$.

(b) Remove one particle from the lattice with $\Delta E_b^{(2)}=-\hbar\omega_L/\pi$ [since this is the total zero-point energy per particle in a long chain as derived in Eq. (2.3)].

(c) Without changing the topology of the lattice the anomalous "spring" is chosen to span the hole so that it is extended over two lattice spacings rather than the usual one per bond. The choice $P'=\frac{1}{2}P$ is made to correspond to a consistent model of the hole.

Then

$$\Delta E_b=\tfrac{1}{2}\hbar\omega_L\pi^{-1}(\tfrac{1}{2}P-3). \qquad (4.19)$$

5. THE DEFECT PAIR

This section is concerned with the interaction of a pair of defects with each other. Several qualitative comments can immediately be made concerning the localized modes. If two defects are very far apart both lead to a localized normal mode of the same frequency. As the defects are brought closer together the degeneracy splits. If the defect coupling is very weak the motion of a given particle located between the two defects can be approximated as a linear combination of the two isolated defect modes. Two independent situations are possible; in the first the motions of the particle of interest due to each defect mode are in phase and in the second they are out of phase. The in-phase motion, which will be called the fitting mode, corresponds to an enhancement of the amplitude of the lattice vibrations while in the out-of-phase motion of the nonfitting case interference occurs. The nonfitting mode has one less node than the fitting one and is associated with the low-frequency member of the split pair.

We now proceed with the detailed analysis of the localized modes of a defect pair.

If a defect is located at lattice point n it induces a displacement from equilibrium at m whose space dependent factor is [generalization of (3.3)]

$$\begin{aligned}u^{(n)}(m)=\gamma^{-1}\{&g(m-n)[(M-M')\omega^2-2(\gamma-\gamma')]\\&+(\gamma-\gamma')[g(m+1-n)+g(m-1-n)]\}u(n)\\&+(1-P)[g(m-n)-g(m-n-1)]u(n+1)\\&+(1-P)[g(m-n)-g(m+1-n)]u(n-1).\end{aligned} \qquad (5.1)$$

Since our differential equations are linear we can apply the superposition principle to find the total displacement which results from defects at $n_1, n_2, \cdots, n_k$:

$$u(m)=\sum_{j=1}^{k}u^{(n_j)}(m). \qquad (5.2)$$

When (5.2) is specialized to a pair of defects, one located at $n_1=0$ and the other at $n_2=n$ the frequencies of normal modes are determined so that the homogeneous equations which relate $u(\pm1)$, $u(0)$, $u(n)$, $u(n\pm1)$ are consistent. These equations are obtained by letting m in (5.2) successively run through $-1, 0, 1, n+1, n, n-1$. The vector u of Eq. (2.17) is

$$u=\{u(-1), u(0), u(1), u(n+1), u(n), u(n-1)\}. \qquad (5.3)$$

The G matrix of Eq. (2.17) has the following form:

$$G=\begin{pmatrix}G_1 & G_2\\ G_2 & G_1\end{pmatrix}, \qquad (5.4a)$$

where

$$G_1=\begin{bmatrix}(1-P)[g(1)-g(0)] & 4f^2(P-Q)g(1) & (1-P)[g(1)-g(2)]\\ (1-P)[g(0)-g(1)] & 4f^2g(0)(P-Q)+(1-P) & (1-P)[g(0)-g(1)]\\ (1-P)[g(1)-g(2)] & 4f^2(P-Q)g(1) & (1-P)[g(1)-g(0)]\end{bmatrix}, \qquad (5.4b)$$

and

$$G_2=\begin{bmatrix}(1-P)[g(n+1)-g(n+2)] & 4f^2(P-Q)g(n+1) & (1-P)[g(n+1)-g(n)]\\ (1-P)[g(n)-g(n+1)] & 4f^2(P-Q)g(n) & (1-P)[g(n)-g(n-1)]\\ (1-P)[g(n-1)-g(n)] & 4f^2(P-Q)g(n-1) & (1-P)[g(n-1)-g(n-2)]\end{bmatrix}. \qquad (5.4c)$$

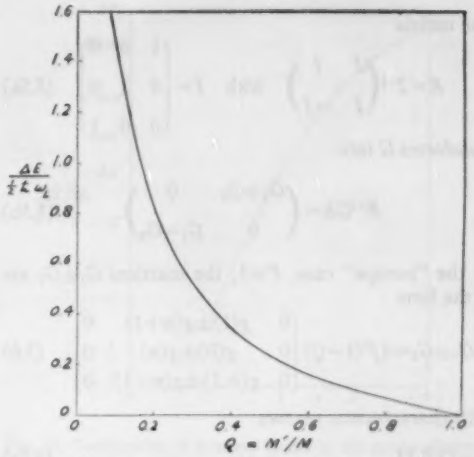


FIG. 9. Self-energy of an isotope mass M' plotted against $Q=M'/M$. ($\frac{1}{2}\hbar\omega_L$ is the zero-point energy corresponding to the maximum frequency of the perfect lattice.)

The matrix

$$R=2^{-\frac{1}{2}}\begin{pmatrix} I & I \\ I & -I \end{pmatrix} \quad \text{with} \quad I=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.5a}$$

transforms G into

$$R^{-1}GR=\begin{pmatrix} G_1+G_2 & 0 \\ 0 & G_1-G_2 \end{pmatrix}. \tag{5.5b}$$

In the "isotope" case, $P=1$, the matrices $G_1\pm G_2$ are of the form

$$G_1\pm G_2=4f^2(1-Q)\begin{bmatrix} 0 & g(1)\pm g(n+1) & 0 \\ 0 & g(0)\pm g(n) & 0 \\ 0 & g(-1)\pm g(n-1) & 0 \end{bmatrix}, \tag{5.6}$$

with characteristic vectors

$$\psi_1=\{1,0,1\}, \tag{5.7a}$$

$$\zeta_1=\frac{1}{2}\left(1,-\frac{2g(1)\pm 2(1-2f^2)g(n)}{g(0)\pm g(n)},1\right)$$

with $\lambda_1=0$, (5.7b)

$$\psi_2=\{1,0,-1\}, \tag{5.8a}$$

$$\zeta_2=\frac{1}{2}\left(1,\frac{\pm[g(n-1)-g(n+1)]}{g(0)\pm g(n)},-1\right)$$

with $\lambda_2=0$, (5.8b)

and

$$\psi_3=\{g(1)\pm g(n+1),\, g(0)\pm g(n),\, g(1)\pm g(n-1)\}, \tag{5.9a}$$

$$\zeta_3=(0,1,0)/[g(0)\pm g(n)], \tag{5.9b}$$

with

$$\lambda_3=4f^2(1-Q)[g(0)\pm g(n)]. \tag{5.9c}$$

The frequency which corresponds to the mode ψ_3 is obtained by letting $\lambda_3=1$ [see equation below (3.15b)]. If we introduce our usual expression (2.8b) for $g(j)$ and f^2 into (5.9c), we have the characteristic equation

$$e^z-(2-Q)/Q=\pm Q^{-1}(1-Q)(-1)^n(e^z+1)e^{-nz}, \quad 0<Q<1. \tag{5.10}$$

This equation has been obtained on the supposition that $n\geq 3$ but it is easy to show that it is also valid when $n=1, 2$. A special analysis is required for these two cases because when $n=2$ the isotopes are next nearest neighbors and when $n=1$ they are nearest neighbors.

Of course in the first approximation as $n\to\infty$, $\exp z=(2-Q)/Q>1$. A series expansion of our required root can be obtained from the Lagrange theorem: Let $f(w)$ and $\psi(w)$ be functions of a complex variable w analytic on and inside a contour c surrounding a point a, and let t be such that the inequality

$$|t\psi(w)|<|w-a| \tag{5.11}$$

is satisfied at all points on the perimeter of C; then the solution

$$\zeta=a+t\psi(\zeta) \tag{5.12}$$

regarded as an equation in ζ, has one root in the interior of C; and further, any function of ζ analytic on and inside C can be expanded as a power series in t by the formula

$$f(\zeta)=f(a)+\sum_{m=1}^{\infty}\frac{t^m}{m!}\frac{d^{m-1}}{da^{m-1}}[f'(a)\{\psi(a)\}^m]. \tag{5.13}$$

In particular the root ζ of (5.12) which lies in C is

$$\zeta=a+\sum_{m=1}^{\infty}\frac{t^m}{m!}\frac{d^{m-1}}{da^{m-1}}[\psi(a)]^m \tag{5.14}$$

If we let $\exp z=\zeta$ then we can apply (5.14) to find

$$e^z=[(2-Q)/Q]\pm 2(-1)^n(1-Q)Q^{-2}[Q/(2-Q)]^n + (2/Q)[(1-Q)/Q]^2 \times\left(\frac{2-Q-2n}{2-Q}\right)\left(\frac{Q}{2-Q}\right)^{2n}+\cdots, \tag{5.15a}$$

and set $f(\zeta)=\frac{1}{4}(\zeta+2+\zeta^{-1})$ in (5.13) to find

$$(\omega/\omega_L)^2=\frac{1}{Q(2-Q)}\pm 2\frac{(-1)^n(1-Q)^2}{Q^2(2-Q)^2}\left(\frac{Q}{2-Q}\right)^n + \frac{(1-Q)^2}{Q^3(2-Q)^3}[4-6Q+3Q^2-4(1-Q)n]\left(\frac{Q}{2-Q}\right)^{2n}+\cdots, \tag{5.15b}$$

so that

$$(\omega/\omega_L)=[Q(2-Q)]^{-\frac{1}{2}}\left\{1\pm\frac{(-1)^n(1-Q)^2}{2(2-Q)}\left(\frac{Q}{2-Q}\right)^n + \frac{1}{2}\frac{(1-Q)^2}{Q^2(2-Q)^2}[3-4Q+2Q^2-4n(1-Q)] \times\left(\frac{Q}{2-Q}\right)^{2n}+\cdots\right\}. \tag{5.16}$$

We see that the frequency of the isolated impurity $f=[Q(2-Q)]^{-\frac{1}{2}}$ is split into two frequencies as two isolated impurities are brought together.

The difference in zero-point energy of the two particle modes associated with a pair of particles n lattice spacings apart and that of two isolated local modes is, as $n\to\infty$

$$\Delta E_D=\tfrac{1}{2}\hbar\sum\Delta\omega = -2n\hbar\omega_L(1-Q)^2[Q/(2-Q)]^{2n}/[Q(2-Q)]^{5/2}, \tag{5.17}$$

since $\frac{1}{2}\hbar[Q(2-Q)]^{-\frac{1}{2}}$ is the zero-point energy per isolated impurity. Hence the localized modes give rise to an attraction between defects. We shall show in the

next section that in three dimensions this attraction is of the Debye or Yukawa type at great distances.

When n is a small integer it is more convenient to solve the characteristic equation (5.10) directly than to employ the Lagrange formula. We have plotted ΔE_D as a function of n in Fig. 10.

The attraction between light isotope defects in a system of a heavy isotopic species leads to a clustering tendency for the light isotope at very low temperatures, and therefore one would expect an ordering into two phases. This separation has been discussed by Prigogine and his collaborators.[8]

The analysis of the interband frequencies of a chain with two isotopic defects is similar to that of the Appendix for the single isotopic impurity. If one impurity is at $-m$ and the other at m in a chain of $2N+1$ atoms (with ends fixed) the equations for the configurational factor of the displacements are

$$M\omega^2 u(n)+\gamma[u(n-1)-2u(n)+u(n+1)] = (M-M')\omega^2\delta(n^2-m^2)u(n), \quad (5.18a)$$

$$u(N)=u(-N)=0, \quad (5.18b)$$

where as usual $\delta(s)=0$ unless $s=0$ in which case it is 1. The solution of this equation is

$$u(n)=\begin{cases} B\sin(N+n)\varphi & \text{if } n<-m \\ A\sin(N-n)\varphi & \text{if } n\geq m \\ Ce^{in\varphi}+De^{-in\varphi} & \text{if } -m\leq n<m \end{cases} \quad (5.19)$$

where

$$M\omega^2=2\gamma(1-\cos\varphi) \quad (5.20)$$

and the parameters A, B, C, D, and φ are determined from the four "connecting" equations which result from successively letting n be $-m-1$, $-m$, m, $m+1$. A bit of elementary algebra yields the four equation relations between A, B, C, D, and φ:

$$(A+B)\sin(N-m)\varphi=2(C+D)\cos m\varphi,$$
$$(A-B)\sin(N-m)\varphi=2i(C-D)\sin m\varphi,$$
$$B\sin(N-m-1)\varphi-A\sin(N-m+1)\varphi +2i(Ce^{-i\varphi}-De^{i\varphi})\sin m\varphi =2(A+B)(1-\cos\varphi)(1-Q)\sin(N-m)\varphi,$$
$$B\sin(N-m-1)\varphi+A\sin(N-m+1)\varphi -2(Ce^{-i\varphi}+De^{i\varphi})\cos m\varphi =2(B-A)(1-\cos\varphi)(1-Q)\sin(N-m)\varphi.$$

The solutions of this set are of two types

(a) The even solutions, $u(n)=u(-n)$, with

$$A=B, \quad C=D, \quad A\sin(N-m)\varphi=2c\cos m\varphi,$$

so that

$$u(n)=A\frac{\sin(N-m)\varphi}{\cos m\varphi}\cos n\varphi, \quad -m\leq n<m$$

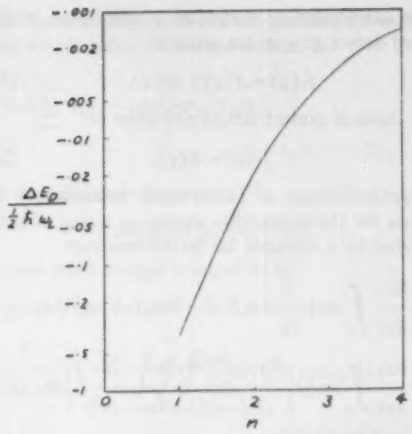


FIG. 10. Contribution of localized modes to the energy of interaction of two isotopes of masses $\frac{1}{2}M$. ($\frac{1}{2}\hbar\omega_L$ is the zero-point energy corresponding to the maximum frequency of the perfect lattice.)

and the characteristic equation for φ is

$$\cot N\varphi=2(Q-1)\tan\tfrac{1}{2}\varphi\cos m\varphi\frac{\sin(N-m)\varphi}{\sin N\varphi}. \quad (5.21)$$

(b) The odd solution, $u(n)=-u(-n)$, with

$$A=-B, \quad C=-D, \quad A\sin(N-m)\varphi=2iC\sin m\varphi,$$

so that

$$u(n)=A\frac{\sin(N-m)\varphi}{\sin m\varphi}\sin n\varphi, \quad -m\leq n<m$$

and the characteristic equation for φ is

$$\sin N\varphi=2(Q-1)\tan\tfrac{1}{2}\varphi\sin(N-m)\varphi\sin m\varphi. \quad (5.22)$$

The characteristic equations which correspond to a single defect at m or $-m$ is

$$\sin N\varphi\cos N\varphi=(Q-1)\tan\tfrac{1}{2}\varphi\sin(N-m)\varphi \times\sin(N+m)\varphi, \quad (5.23)$$

which when $m=0$ becomes

$$\sin N\varphi[\cot N\varphi-(Q-1)\tan\tfrac{1}{2}\varphi]=0$$

which is equivalent to (A-7) and (A-8) in the Appendix.

If we combine (5.22) and (5.21) we find that the values of φ which lead to normal mode frequencies are zeros of the function

$$f(\varphi)=\{\cos N\varphi+2\epsilon\tan\tfrac{1}{2}\varphi\cos m\varphi\sin(N-m)\varphi\} \times\{\sin N\varphi+2\epsilon\tan\tfrac{1}{2}\varphi\sin(N-m)\varphi\sin m\varphi\} =A(\varphi)+2\epsilon B(\varphi)+\epsilon^2C(\varphi), \quad (5.24a)$$

where

$$\epsilon=(1-Q), \quad (5.24b)$$

$$A(\varphi)=\cos N\varphi\sin N\varphi=\tfrac{1}{2}\sin 2N\varphi, \quad (5.24c)$$

$$B(\varphi)=\tfrac{1}{2}\tan\tfrac{1}{2}\varphi\{\cos 2m\varphi-\cos 2N\varphi\}, \quad (5.24d)$$

$$C(\varphi)=2\tan^2\tfrac{1}{2}\varphi\sin^2(N-m)\varphi\sin 2m\varphi. \quad (5.24e)$$

The corresponding values of φ associated with an isolated defect at $\pm m$ are zeros of

$$f_0(\varphi)=A(\varphi)+\epsilon\beta(\varphi), \tag{5.25}$$

while those of perfect lattice are zeros of

$$f_1(\varphi)=A(\varphi). \tag{5.26}$$

A generalization of (4.12) leads immediately to a formula for the interaction energy of a pair of defects separated by a distance $2m$ lattice spacings

$$\Delta E=\frac{\hbar\omega_L}{4\pi i}\int_C \sin\tfrac{1}{2}z\frac{d}{dz}\{\log f(\varphi)-2\log f_0+\log f_1\}dz$$

$$=\frac{\hbar\omega_L}{4\pi i}\int_C \sin\tfrac{1}{2}z\left\{\frac{\beta'-[\alpha^2]'+\epsilon(\alpha\beta'-2\beta\alpha')}{(1+\epsilon\alpha)(1+2\epsilon\alpha+\beta\epsilon^2)}\right\}dz, \tag{5.27a}$$

where

$$\alpha=B/A=\tan\tfrac{1}{2}z\left\{\frac{\cos 2mz-\cos 2Nz}{\sin 2Nz}\right\}, \tag{5.27b}$$

$$\beta=C/A=4\frac{\tan^2\tfrac{1}{2}z\sin^2(N-m)z\sin 2mz}{\sin 2Nz}. \tag{5.27c}$$

The contour C is again around the counter clockwise rectangle given in Fig. 8.

It can be shown that in the limit as $N\to\infty$, the integrand is an even function of y along the line $z=iy$ and an odd function along $z=\pi+iy$. Hence the integral from $-\pi+iy$ to $\pi+iy$ vanishes. It can also be shown that as $N\to\infty$, the values of α and β along $z=x+ia$ are

$$\alpha\sim i\tan\tfrac{1}{2}(x+ia)+O(\exp -Na), \tag{5.28a}$$

$$\beta\sim-\tan^2\tfrac{1}{2}(x+ia)+O(\exp -Na). \tag{5.28b}$$

Hence the integrand of (5.27a) along $z=x+ia$ is of $O(\exp -Na)$ and can be neglected. One also finds the integral along $z=x-ia$ to vanish in the same way. Since, as $N\to\infty$

$$\alpha=-\tanh\tfrac{1}{2}y,\quad \beta=(\tanh^2\tfrac{1}{2}y)(1-\exp[-4my]),$$

we finally obtain, after letting $a\to\infty$,

$$\Delta E=-\frac{\epsilon^2\hbar\omega_L}{2\pi}\int_0^\infty\frac{[\sinh\tfrac{1}{2}y\tanh\tfrac{1}{2}y]e^{-4my}}{(1-\epsilon\tanh\tfrac{1}{2}y)}$$

$$\times\left\{\frac{4m\tanh\tfrac{1}{2}y(1-\epsilon\tanh\tfrac{1}{2}y)-\operatorname{sech}^2\tfrac{1}{2}y}{1-2\epsilon\tanh\tfrac{1}{2}y+\epsilon^2\tanh^2\tfrac{1}{2}y[1-\exp(-4my)]}\right\}dy. \tag{5.29}$$

The authors have not been able to evaluate this integral as a function of ϵ and m. However it is easily calculated as $\epsilon\to 0$ and as $\epsilon\to 1$.

The energy of interaction of two weak isotopic defects separated by $2m$ lattice spacings is found by setting $\epsilon=0$ in the integrand of (5.29):

$$\Delta E=\frac{\hbar\omega_L\epsilon^2}{2\pi}\int_0^\infty\sinh\tfrac{1}{2}y\,d\{\tanh^2\tfrac{1}{2}y\exp -4my\} \tag{5.30}$$

$$=-\frac{\epsilon^2\hbar\omega_L}{4\pi}\int_0^\infty e^{-4my}\tanh^2\tfrac{1}{2}y\cosh\tfrac{1}{2}y\,dy$$

$$=-\frac{\epsilon^2\hbar\omega_L}{4\pi}\left\{\frac{4m}{(4m)^2-\frac{1}{4}}-\int_0^\infty\operatorname{sech}\tfrac{1}{2}y\exp(-4my)dy\right\} \tag{5.31}$$

$$=-\frac{\epsilon^2\hbar\omega_L}{4\pi}\left\{\frac{4m}{(4m)^2-\frac{1}{4}}-\psi(\tfrac{3}{4}+2m)+\psi(\tfrac{1}{4}+2m)\right\},$$

where as usual

$$\psi(z)=\frac{d}{dz}\log\Gamma(z)=-\gamma-\frac{1}{z}+\sum_{n=1}^{\infty}\frac{z}{n(z+n)},$$

with $\gamma=0.57721$. The asymptotic formula for large z

$$\psi(z)=\log z-(2z)^{-1}-\sum_{n=1}^{m}B_{2n}z^{-2n}/2z+O(z^{-2m-2}),$$

and Nörlands factorial expansion

$$\psi(a+z)-\psi(a)=\frac{z}{a}-\frac{1}{2}\frac{z(z-1)}{a(a+1)}+\cdots$$

are useful for finding the interaction of two widely separated isotopic defects. A finite series for ΔE is

$$\Delta E=-\frac{\epsilon^2\hbar\omega_L}{\pi}\left\{\frac{4m}{(8m)^2-1}-\frac{\pi}{4}+\left[1-\tfrac{1}{3}+\cdots-\frac{1}{8m-1}\right]\right\}, \tag{5.32}$$

while the first few terms of the asymptotic series for large separations are

$$\Delta E\sim-\frac{(Q-1)^2\hbar\omega_L}{\pi}\left\{\frac{1}{(8m)^3}-\frac{2}{(8m)^5}+\cdots\right\}. \tag{5.33}$$

It is remarkable that the asymptotic formula is valid over the entire range of integral values of m. When $m=1$ it is found that the exact equation (5.32) yields $\Delta E=-1.903\times10^{-3}\hbar\omega_L\epsilon^2/\pi$ while the first two terms of the asymptotic formula yield $E=-1.89\times10^{-3}\hbar\omega_L\epsilon^2/\pi$ with even better agreement when $m\geq 2$.

The result (5.32) has also been obtained from second-order perturbation theory.

In the limit of a strong defect ($Q=0$) we use the fact that $\epsilon=1$ so that

$$1-\epsilon\tanh\tfrac{1}{2}y=\operatorname{sech}\tfrac{1}{2}y\exp -\tfrac{1}{2}y,$$

which yields

$$\Delta E=-\frac{\epsilon^2\hbar\omega_L}{2\pi}\int_0^\infty \frac{\sinh^2\tfrac{1}{2}y e^{-(4m-1)y}\{4m\sinh\tfrac{1}{2}y-e^{\frac{1}{2}y}\}dy}{1-\sinh^2\tfrac{1}{2}y\exp\{-(4m-1)y\}}. \tag{5.34}$$

With the possible exception of the choice $m=1$, the term $\sinh^2\frac{1}{2}y\exp[-(4m-1)y]$ in the denominator of the integrand can be neglected when compared with 1. Hence the integration becomes elementary and yields

$$\Delta E=-\frac{2\epsilon^2\hbar\omega_L(4m-1)}{\pi(8m-5)(8m-3)(8m-1)(8m+1)}. \tag{5.35}$$

Corrections for small m are easily obtained by expanding the denominator in powers of $\sinh^2\frac{1}{2}y\exp[-(4m-1)y]$.

The first few terms in an asymptotic series for ΔE which is valid for all ϵ when m is large is obtained by expanding all terms in (5.29) as a power series in y. Then

$$\Delta E\sim-\frac{\epsilon^2\hbar\omega_L}{\pi}\frac{1}{(8m)^2}\left\{1+\frac{3}{4m}\epsilon+O(m^{-2})\right\}, \tag{5.36}$$

which is consistent with (5.35) and (5.33). It is to be noted that the first-order term has the same form for all values of $\epsilon=1-Q$.

Note added in proof.—In the case of a pair of holes Q is to be replaced by $2P^{-1}$. If one mass defect corresponds to $\epsilon_1=(1-Q_1)$ and the other to $\epsilon_2=(1-Q_2)$ the first-order term in (5.36) has ϵ^2 replaced by $\epsilon_1\epsilon_2$. Hence the interaction between a heavy and a light defect is repulsive.

6. DEFECTS IN THREE-DIMENSIONAL LATTICES

We now sketch the generalization of the discussion of previous sections to three-dimensional lattices.

The normal frequencies of the perfect three-dimensional simple cubic lattice are defined by

$$M\omega^2=2\sum_{p=1}^{3}\gamma_p\{1-\cos(2\pi r_p/N_p)\};$$
$$r_p=1,2,\cdots,N_p, \tag{6.1}$$

with the maximum frequency

$$M\omega_L^2=4(\gamma_1+\gamma_2+\gamma_3). \tag{6.2}$$

The zero-point energy is found to be

$$E=\tfrac{1}{2}\hbar\Sigma\omega$$
$$=\frac{\hbar}{(2m)^{\frac{1}{2}}}\left\{\prod_{p=1}^{3}\frac{N_p}{2\pi}\right\}\int\int\int_0^{2\pi}\{\sum_{p=1}^{3}\gamma_p(1-\cos\varphi_p)\}^{\frac{1}{2}}$$
$$\times d\varphi_1 d\varphi_2 d\varphi_3. \tag{6.3}$$

If one component of the displacement of the mass at lattice point j_1, j_2, j_3 is represented by $u(j_1,j_2,j_3)$ then the operator L which is a generalization of that given by (2.5) is

$$\begin{aligned}Lu(j_1,j_2,j_3)&=M\omega^2u(j_1,j_2,j_3)+\gamma_1[u(j_1+1,j_2,j_3)\\&\quad-2u(j_1,j_2,j_3)+u(j_1-1,j_2,j_3)]\\&\quad+\gamma_2[u(j_1,j_2+1,j_3)-2u(j_1,j_2,j_3)\\&\quad+u(j_1,j_2-1,j_3)]+\gamma_3[u(j_1,j_2,j_3+1)\\&\quad-2u(j_1,j_2,j_3)+u(j_1,j_2,j_3-1)].\end{aligned} \tag{6.4}$$

The Green's function satisfying

$$Lg(j_1,j_2,j_3)=\begin{cases}\gamma_1+\gamma_2+\gamma_3 & \text{if } j_1=j_2=j_3=0\\ 0 & \text{otherwise}\end{cases} \tag{6.5a}$$

is readily verified to be

$$g(j_1,j_2,j_3)=-\frac{\gamma_1+\gamma_2+\gamma_3}{2\pi^3}\int\int\int_0^{\pi}\frac{\cos j_1\varphi_1\cos j_2\varphi_2\cos j_3\varphi_3}{\alpha(\gamma_1+\gamma_2+\gamma_3)-\gamma_1\cos\varphi_1-\gamma_2\cos\varphi_2-\gamma_3\cos\varphi_3}d\varphi_1 d\varphi_2 d\varphi_3, \tag{6.5b}$$

where

$$\alpha=1-\tfrac{1}{2}M\omega^2(\gamma_1+\gamma_2+\gamma_3)^{-1} \tag{6.5c}$$
$$=1-2f^2. \tag{6.5d}$$

An extensive program for the computation of this Green's function is being carried out. An interesting asymptotic formula is

$$g\sim-\frac{\gamma_1+\gamma_2+\gamma_3}{(\gamma_1\gamma_2\gamma_3)^{\frac{1}{2}}}\frac{1}{4\pi R}\exp\{-2f(\gamma_1+\gamma_2+\gamma_3)^{\frac{1}{2}}R\} \tag{6.6}$$

valid for large $R=(j_1^2\gamma_1+j_2^2\gamma_2+j_3^2\gamma_3)^{\frac{1}{2}}$. This Green's function has also been discussed by Koster and Slater.[1]

As for the linear lattice, the mode corresponding to a single isotope of mass $M'=QM$ is a symmetric vibration with a discrete frequency which is the solution of the transcendental equation

$$\omega_0^2(M-M')g(0,0,0)=\gamma_1+\gamma_2+\gamma_3, \tag{6.7a}$$

or

$$4f_0^2(1-Q)g(0,0,0)=1. \tag{6.7b}$$

The modes corresponding to a single hole in the lattice are the three antisymmetric pulsating ones with discrete frequency levels given by

$$(\gamma_1-\gamma_1')[g(2,0,0)-g(0,0,0)]=\gamma_1+\gamma_2+\gamma_3, \tag{6.8a}$$
$$(\gamma_2-\gamma_2')[g(0,2,0)-g(0,0,0)]=\gamma_1+\gamma_2+\gamma_3, \tag{6.8b}$$
$$(\gamma_3-\gamma_3')[g(0,0,2)-g(0,0,0)]=\gamma_1+\gamma_2+\gamma_3. \tag{6.8c}$$

The interstitial, placed at the center of a cell of the lattice, is topologically different from a lattice point

as it has eight nearest neighbors. If the force constant for the interactions with these nearest neighbors is taken as γ', then there are four antisymmetric modes with discrete frequencies satisfying

$$\gamma'\{g(0,0,0)-g(1,1,1)+g(1,0,0)-g(0,1,1)+g(0,1,0)-g(1,0,1)+g(1,1,0)-g(0,0,1)\}=\gamma_1+\gamma_2+\gamma_3, \quad (6.9a)$$

$$\gamma'\{g(0,0,0)-g(1,1,1)+g(1,0,0)-g(0,1,1)-g(0,1,0)+g(1,0,1)-g(1,1,0)+g(0,0,1)\}=\gamma_1+\gamma_2+\gamma_3, \quad (6.9b)$$

$$\gamma'\{g(0,0,0)-g(1,1,1)-g(1,0,0)+g(0,1,1)+g(0,1,0)-g(1,0,1)-g(1,1,0)+g(0,0,1)\}=\gamma_1+\gamma_2+\gamma_3, \quad (6.9c)$$

$$\gamma'\{g(0,0,0)-g(1,1,1)-g(1,0,0)+g(0,1,1)-g(0,1,0)+g(1,0,1)+g(1,1,0)-g(0,1,1)\}=\gamma_1+\gamma_2+\gamma_3. \quad (6.9d)$$

If two isotopes occupy lattice sites (0,0,0) and (n_1,n_2,n_3) then, as for the linear lattice, fitting and antifitting modes are possible with normal frequencies satisfying

$$4f^2(1-Q)\{g(0,0,0)\pm g(n_1,n_2,n_3)\}=1. \quad (6.10a)$$

The solutions of this equation give a splitting about the solution for the case of one isotope. The contribution of localized modes to the energy of interaction of the isotopes is estimated from the second term in the Lagrange expansion of f, using (6.10a) in the form

$$\frac{1}{4f^2g(0,0,0)}=1-Q\pm(1-Q)\frac{g(n_1,n_2,n_3)}{g(0,0,0)}. \quad (6.10b)$$

If

$$\zeta^{-1}=4f^2g(0,0,0), \quad (6.11a)$$

$$f(\zeta)=\omega/\omega_L, \quad (6.11b)$$

and

$$\varphi(\zeta)=g(n_1,n_2,n_3)/g(0,0,0), \quad (6.11c)$$

then the second term in the Lagrange expansion of $f(\zeta)$ about $\zeta=1-Q$ (i.e., $f=f_0-\omega/\omega_L$) is

$$\tfrac{1}{2}(1-Q)^2\left[\frac{d}{d\zeta}\{f'(\zeta)\varphi^2(\zeta)\}\right]_{\zeta=1-Q}, \quad (6.12a)$$

which for large n_1, n_2, n_3 becomes

$$\tfrac{1}{2}(1-Q)^2\left[f'(\zeta)\frac{d}{d\zeta}\varphi^2(\zeta)\right]_{\zeta=(1-Q)} \quad (6.12b)$$

$$=\tfrac{1}{2}(1-Q)^2\left[\{f'(\zeta)\}^2\frac{d}{df}\varphi^2(f)\right]_{f=f_0}, \quad (6.12c)$$

$$\sim A\left[\frac{d}{df}\frac{1}{R^2}\exp\{-4f(\gamma_1+\gamma_2+\gamma_3)^{\frac{1}{2}}R\}\right]_{f=f_0}, \quad (6.12d)$$

with

$$R^2=n_1^2\gamma_1+n_2^2\gamma_2+n_3^2\gamma_3; \quad (6.12e)$$

and where A is a positive constant and use has been made of the asymptotic formula (6.6a). The energy of interaction due to localized modes is hence given by

$$V(R)\sim -BR^{-1}\exp -CR, \quad (6.13)$$

where B and C are positive constants (i.e., functions of Q only and independent of n_1, n_2, n_3). The force has the sign of an attraction.

In the case of an isotropic discrete quantum field $\gamma_1=\gamma_2=\gamma_3=\gamma$ so that $R^2=\gamma(n_1^2+n_2^2+n_3^2)$ and the interaction $V(r)$ due to localized modes between our pair of defects is of the Yukawa type at large distances.

Since a crystal lattice corresponds to $\gamma_1\gg\gamma_2=\gamma_3$ the force law (6.13) is not spherically symmetrical and depends on the orientation with respect to crystal axes of the line which connects the two defects.

A more detailed discussion of the three-dimensional lattice will be given in a later publication.

APPENDIX. THE ISOLATED ISOTOPIC IMPURITY IN A LINEAR CHAIN

We shall give a detailed analysis of the "isotopic" impurity in this section. It will serve as an example of how localized modes make their appearance in an exact treatment of the modes in the band. A similar discussion for the electron theory of solids has been given by Koster and Slater,[1] and by Saxon and Hutner.[2]

The equations for the time-independent factor of the displacements are

$$M\omega^2u(n)+\gamma[u(n-1)-2u(n)+u(n+1)]=0, \quad n\neq 0, \quad (A.1a)$$

$$M'\omega^2u(0)+\gamma[u(-1)-2u(0)+u(1)]=0, \quad (A.1b)$$

$$u(N)=u(-N)=0, \quad (A.1c)$$

when one deals with a system of $2N$ particles of mass M and one of mass M' located in the middle of the chain while the end particles are held fixed. One finds

$$u(n)=\begin{cases}A\sin(N-n)\varphi & \text{if } n\geq 0\\ B\sin(N+n)\varphi & \text{if } n<0\end{cases} \quad (A.2)$$

while

$$M\omega^2=2\gamma(1-\cos\varphi). \quad (A.3)$$

One relates the constants A and B by substituting (A.2) into the connecting equations of (A.1) which correspond to $n=0$ and $n=-1$. One finally obtains the following relations between A, B, and φ:

$$B\sin(N-1)\varphi-A\sin(N+1)\varphi=2A(1-Q)(1-\cos\varphi)\sin N\varphi, \quad (A.4)$$

$$(A-B)\sin N\varphi=0. \quad (A.5)$$

Two possible situations arise

$$\text{(a)}\quad A=B \quad (A.6)$$

or

$$\text{(b)}\quad \sin N\varphi=0. \quad (A.7)$$

Condition (b) leads to the odd solutions with $A=-B$ while (a) leads to the even solutions. Since the possible values of φ in the odd solutions are exactly the same as those in the perfect lattice, $\{2\pi j/2N\}$, no shift occurs in the corresponding frequencies.

The even solutions of (A.1) have the characteristic equation

$$\cot N\varphi=(Q-1)\tan\tfrac{1}{2}\varphi, \tag{A.8}$$

for the determination of the φ's. The solutions of (A.8) can be determined by plotting (see Fig. 11) both the left and right hand sides of (A.8) on the same graph and locating their intersection points.

When $Q=1$ the right-hand side of (A.8) vanishes identically so that the possible values of φ are the intersections of the various branches of $\cot N\varphi$ with the φ axis, $\varphi_j=(2j-1)\pi/2N$. When $Q\neq 1$ one intersection occurs in each interval $\{\pi j/N, \pi(j+1)/N\}$ except for the interval $[(N-1)\pi/N, \pi]$ when $Q<1$. No intersection occurs in this interval because the function $(Q-1)\tan\frac{1}{2}\varphi$ always lies below $\cot N\varphi$ in the entire interval.

The introduction of a heavy impurity, $Q>1$, causes a displacement of the various intersections to the left and yields an associated decrease in normal mode frequencies (the largest possible displacement being $\pi/2N$); that of a light impurity $Q<1$ causes displacements to the right with associated increases in normal mode frequencies (with the exception of the "lost mode" the displacement is limited by $\pi/2N$). The lost mode in the light impurity case corresponds to the localized mode which emerges from the band. It is resurrected by assuming that $\varphi=\pi+i\vartheta$ and considering the characteristic equation for ϑ, $\coth N\vartheta=(Q-1)\times\coth\frac{1}{2}\vartheta$. Then in the limit as $N\to\infty$

$$\exp\vartheta=Q/(Q-2), \tag{A.9}$$

while

$$(\omega/\omega_L)^2=\tfrac{1}{2}(1-\cosh z)=1/Q(2-Q), \tag{A.10}$$

which is exactly (3.20).

The change in zero-point energy of the lattice with one isotope has been worked out in detail in Sec. 4; we shall now check this result for the limiting case $Q\to 0$. In this limit, Eq. (A.8) becomes

$$\cot N\varphi=-\tan\tfrac{1}{2}\varphi, \tag{A.11}$$

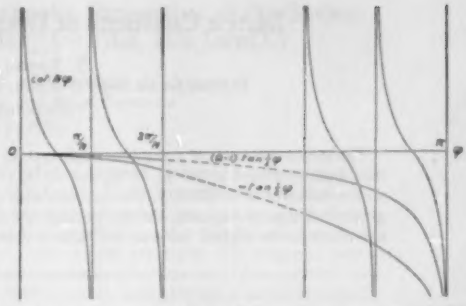


FIG. 11. The solution of Eq. (A.8) can be obtained graphically from the curves sketched above. The right-hand side of (A.8) is sketched for a value $0<Q<1$ and for $Q=0$.

which can be solved exactly:

$$\varphi_j=[(2j-1)\pi/(2N-1)];\ j=1, 2, \cdots, N-1. \tag{A.12}$$

As can be seen in Fig. 11, the first solution

$$\varphi_1=\pi/(2N-1)$$

is near $\pi/2N$ whereas $\varphi_{N-1}=(2N-3)\pi/(2N-1)$ is near $(2N-2)\pi/2N$. The change in zero-point energy contributed by the shifts of the levels in the band is, therefore,

$$\tfrac{1}{2}\hbar\omega_L\left\{\sum_{j=1}^{N-1}\sin\frac{(2j-1)\pi}{2(2N-1)}-\sum_{j=1}^{N}\sin\frac{(2j-1)\pi}{4N}\right\} \tag{A.13}$$

$$=\frac{1}{2}\hbar\omega_L\left\{\frac{1}{2}\operatorname{cosec}\frac{\pi}{2(2N-1)}-\frac{1}{2}-\frac{1}{2}\operatorname{cosec}\frac{\pi}{4N}\right\} \tag{A.14}$$

$$\simeq\frac{1}{2}\hbar\omega_L\left(-\frac{1}{2}-\frac{1}{\pi}\right) \tag{A.15}$$

for large N.

Since the level out of the band contributes

$$\tfrac{1}{2}\hbar\omega_L[Q(2-Q)]^{-\frac{1}{2}}$$

as given by Eq. (4.7a), the change in zero-point energy is thus

$$\frac{1}{2}\hbar\omega_L\left\{[Q(2-Q)]^{-\frac{1}{2}}-\frac{1}{2}-\frac{1}{\pi}+O(Q^{\frac{1}{2}})\right\}. \tag{A.16}$$

This agrees with the expansion given in Eq. (4.15).

Lattice Constants of Graphite at Low Temperatures*

Y. Baskin and L. Meyer
Institute for the Study of Metals, University of Chicago, Chicago, Illinois
(Received May 2, 1955)

The lattice constants of a graphite single crystal, natural graphite powder, and artificial graphite powder have been measured at 297°K, 78°K, and 4.2°K. The a-spacing does not change in this temperature range within the error of ±0.0005 A. The material with the greatest c-spacing showed between 297° and 78° the greatest change of c-spacing and the smallest change below 78°. The c-spacing of the artificial powder did not return to its original value at 297° after a thermal cycle, demonstrating irreversible deformation.

THE measurements at low temperatures of the physical properties of graphite such as electrical conductivity are usually slightly irreproducible.[1,2] Carbon thermometers used in the liquid helium temperature range generally have to be recalibrated each time they are used. It was suspected that this effect is due to the unusually large anisotropy of the thermal expansion of graphite which can cause mechanical deformation in every thermal cycle. In order to test this hypothesis precision measurements were made of the lattice constants of a single crystal of graphite (from Ticonderoga, New York), spectroscopically pure natural graphite powder (National Carbon Corporation No. SP 1) and spectroscopically pure artificial graphite (National Carbon Corporation No. SP 2) at room temperature (297°K), liquid nitrogen temperature (78°K), and liquid helium temperature (4.2°K). The measurements were performed in the low-temperature precision x-ray spectrometer developed in this laboratory and described elsewhere.[3]

The c-axis spacing was derived by the method of Nelson and Riley[4] from the maxima of the 002, 004, 006, and 008 peaks produced by Ni-filtered Cu radiation; each peak was scanned at intervals of one minute of arc. The a-axis spacing was derived from the 110 peak only. The accuracy with which the peaks were determined is ±0.005° yielding an accuracy for the spacing of ±0.0005 A. The temperature was known to ±0.3°K. Table I contains a summary of the results for the c-spacing.

Table I. Interlayer spacing of graphite as a function of temperature.

T°K	Single crystal A	Natural power SP 1 A	Artificial power SP 2 A
297°	3.3538	3.3542	3.3600
78°	3.3378	3.3378	3.3392
4.2°	3.3360	3.3367	3.3388
297° after cooling	3.3538	3.3542	3.3590
Expansion coefficient 78°–297°K	$7.34\cdot10^{-5}$	$7.59\cdot10^{-5}$	$9.52\cdot10^{-5}$

The change of interlayer spacing with temperature depends strongly on the nature of the sample: The material with the largest interlayer spacing showed the greatest change of spacing, between 297° and 78° and the smallest change below 78°. A large c-spacing is usually considered to be a sign of less perfect crystallites.[5] Since anisotropic changes of volume, especially in polycrystalline material, are often accompanied by plastic deformation, it does not seem surprising that a temperature exists at which the thermal forces trying to contract the specimen are balanced by the opposing mechanical strain from surrounding crystallites, and that this temperature depends on the size and the quality of the crystals. At the lowest temperatures the single crystal which had the smallest fraction of imperfections among the specimens showed the greatest amount of contraction with the lowering of the temperature.

The irreversible changes possible during a thermal cycle are demonstrated by the fact that the c-spacing of the artificial graphite did not return to its original value at 297°K after being cooled to 4.2°K.

As Walker *et al.*[6] found a practically linear variation of interlayer spacing with temperature down to 78°K, an average thermal expansion coefficient α defined by

$$d_T = d_{T_0} + \alpha(T - T_0),$$

d_T being the interlayer spacing at the temperature T, and calculated between 297° and 78°, is given in the bottom row of Table I. The value of α for the artificial powder agrees with that found by Walker *et al.*; the α values for the natural powder and the single crystal are distinctly lower.

The a-axis spacing was found to be 2.4589±0.0005 A at 297°K, changing less than 0.0005 A when the specimen was cooled to 78°K.

* Supported in part by U. S. Atomic Energy Commission.
[1] Meyer, Picus, and Johnston, National Bureau of Standards Circular No. 519, October, 1952, p. 249.
[2] W. Primak and L. H. Fuchs, Phys. Rev. **95**, 22 (1954).
[3] Barrett, Long, Meyer, and Walker (to be published).
[4] J. B. Nelson and D. P. Riley, Proc. Phys. Soc. (London) **57**, 477 (1945).
[5] R. E. Franklin, Acta Cryst. **4**, 253 (1951).
[6] Walker, McKinstry, and Wright, Ind. Eng. Chem. **45**, 1711 (1953).

Neutron Diffraction Study of the Magnetic Properties of the Series of Perovskite-Type Compounds $[(1-x)\text{La}, x\text{Ca}]\text{MnO}_3$†

E. O. Wollan and W. C. Koehler
Oak Ridge National Laboratory, Oak Ridge, Tennessee
(Received May 9, 1955)

A study has been made of the magnetic properties of the series of perovskite-type compounds $[(1-x)\text{La}, x\text{Ca}]\text{MnO}_3$. The investigations have been made primarily by neutron diffraction methods, but x-ray diffraction measurements of lattice distortions and ferromagnetic saturation data are also included. This series of compounds exhibits ferromagnetic and antiferromagnetic properties which depend upon the relative trivalent and tetravalent manganese ion content. The samples are purely ferromagnetic over a relatively narrow range of composition ($x \sim 0.35$) and show simultaneous occurrence of ferromagnetic and antiferromagnetic phases in the ranges ($0 < x < 0.25$) and ($0.40 < x < 0.5$). Several types of antiferromagnetic structures at $x = 0$ and $x > 0.5$ have also been determined. The growth and mixing of the various phases have been followed over the whole composition range, the ferromagnetic and antiferromagnetic moment contributions to the coherent reflections have been determined, and Curie and Néel temperatures have been measured. The results have been organized into a scheme of structures and structure transitions which is in remarkable accord with Goodenough's predictions based on a theory of semicovalent exchange.

1. INTRODUCTION

A NEUTRON diffraction study has been made of the magnetic properties of the series of compounds $[x\text{Ca}, (1-x)\text{La}]\text{MnO}_3$, where x varies from zero to unity. For all values of the parameter x, these mixed oxides crystallize in various modifications of the perovskite structure. The ideal cube of a perovskite contains one molecule ABO_3, where A is a large ion, for example, La^{+3}, Nd^{+3}, Ca^{+2}, Sr^{+2}, Ba^{+2}, etc., located at the cube center; B is a small ion, Mn^{+3}, Co^{+3}, Ti^{+4}, Mn^{+4}, etc., situated at cube corners; and the three oxygen ions are at the midpoints of the cube edges. Many examples of the perovskite structure are known, only a few of which are perfectly cubic.

Jonker and Van Santen[1] have investigated the magnetic properties of a number of these perovskite-type compounds and have found that certain of the mixed oxides containing manganese or cobalt become ferromagnetic with Curie points ranging downward from about room temperature. Their measurements of saturation moments and Curie temperatures for the $(\text{La}, \text{Ca})\text{MnO}_3$ series to be considered here are reproduced together with associated data of this report in Figs. 12 and 15 of Secs. 9 and 10. Their saturation measurements represent the average ferromagnetic moments of the macroscopic specimens and the Curie points refer to the temperature of transition to the ferromagnetic state. Application of the neutron-diffraction technique has made it possible to obtain more detailed information about the magnetic properties of these compounds. The ferromagnetic properties have been more fully investigated and the system has been shown also to have a variety of antiferromagnetic structures.

2. PREPARATIONS AND CHEMICAL ANALYSIS OF THE SAMPLES

The manganite samples were prepared by the ceramic technique outlined by Jonker and Van Santen.[1] Oxides, carbonates or hydroxides of lanthanum, calcium and manganese in the desired proportions were mixed under ethanol in a ball mill, dried at 100°C, and prefired, usually in air, at 900°–1000°C.[2] The products resulting from prefiring were mixed again, pressed without binder into briquets and finally fired in a resistance furnace at temperatures ranging from 1100°C to 1400°C. A few preparations were also made from solutions of the nitrates of the metals.

Diffraction samples for many preparations were prepared in the usual way. The sintered briquets were broken up to a fine powder which was packed into a thin-walled aluminum or vanadium cylindrical cell. For preparations to be studied as a function of applied magnetic field, the briquets themselves were used.

The ionic composition of the preparations depends markedly upon the temperature and atmosphere of firing. This has been pointed out by Jonker and Van Santen and has also been observed in the present investigations. It has been possible, for example, to obtain LaMnO_3 preparations with tetravalent manganese concentrations ranging from 2% to 35%: the lowest concentration resulted from firing in nitrogen at 1400°C, the highest from firing in oxygen at 1100°C. Similarly, preparations of CaMnO_3 containing 71, 80, 91, and 97% tetravalent manganese were found after firing in oxygen at 1400°C, 1350°C, 1300°C, and 1075°C, respectively.

Because the magnetic properties of these mixed manganites are so strongly dependent on the percent Mn^{+4} content, this quantity has in all cases been de-

† A preliminary account of this work was presented at the Conference on Ferrimagnetism, U. S. Naval Ordnance Laboratory, Washington, D. C., October 11–12, 1954.

[1] G. H. Jonker and J. H. Van Santen, Physica **16**, 337 (1950); **19**, 120 (1953).

[2] We were aided in the preparation of starting materials by Mr. D. E. LaValle of the Analytical Chemistry Division.

TABLE I. Data summary.

Sample	No.	% Mn^{+4}	Temp.[a]	θ_c°K	θ_n°K	Structure type	μ_A^2	μ_B^2	$\bar{\mu}_{sat}$
$LaMnO_3$	12	2	*He, N, R, C, B*		100	*A*	15.12		
$LaMnO_3$	34	9	*H, R*			*A, B*	13.68	1.36	0.49
$LaMnO_3$	1	10	*He, N, R, W*		140	*A, B*	13.24		0.53
$LaMnO_3$	65	14	*He, N, R*			*A, B*	11.68	3.48[b]	0.92
$LaMnO_3$	73	18	*He, P, N, R*	170		*B, A*	3.44	8.92[c]	
$LaMnO_3$	3	20	*H, P, N, R, W*	170	160	*B, A*	2.08	11.44[d]	2.86
(0.89 La−0.11 Ca)MnO_3	61	29	*He, R*			*B*[e]			
(0.85 La−0.15 Ca)MnO_3	21	38	*N, R, W*	190		*B, A*	0.4	13.80	
(0.85 La−0.15 Ca)MnO_3	23	18	*N, R, W*	170		*B, A*	0.44	12.00	2.22
(0.75 La−0.25 Ca)MnO_3	28	32	*N, R, W*	180		*B*		12.60	3.48
(0.65 La−0.35 Ca)MnO_3	70	35	*P, N, R*	250		*B*		13.24[f]	
(0.6 La−0.4 Ca)MnO_3	4	44	*P, N, R, W*	260		*B*[g]		10.60	3.47
(0.5 La−0.5 Ca)MnO_3	16	49	*N, R*	260		*B*[g]		10.64	3.19
(0.5 La−0.5 Ca)MnO_3	41	51	*N, R, W*	250		*B*[g]		10.04	
(0.5 La−0.5 Ca)MnO_3	75	46	*He, N, R*			*B*[g,h]		9.96	
(0.5 La−0.5 Ca)MnO_3	66	58	*He, R*			*CE*			
(0.4 La−0.6 Ca)MnO_3	5	71	*H, P, N, R, W*		170	*CE*			
(0.25 La−0.75 Ca)MnO_3	18	78	*N, R, W*		165	*CE*			0.18
(0.20 La−0.80 Ca)MnO_3	74	80	*P, N, R*		165	*C*	8.92		
$CaMnO_3$	6	71	*H, N, R, W*		130	?			
$CaMnO_3$	20	80	*P, N, R*		130	*G*	3.40		
$CaMnO_3$	26	91	*N, R, W*		130	*G*	5.92		
$CaMnO_3$	32	97	*P, N, R*		110	*G*	7.04		

[a] The temperatures have been coded to save space: *He* = 4.2°K, *H* = 20.4°K, *P* = 46°K, *N* = 77°K, *R* = 293°K, *C* = 198°K, *B* = 370°K, and *W* indicates that part of the pattern was observed at successively higher temperature values in the course of sample warmup.
[b] No field dependence for either type of reflection was observed at 5000 oersteds.
[c] At 4.2°K the 200 difference reflection was reduced to 0.41 of its no field value by 5000 oersteds.
[d] At 77°K the 200 difference reflection was reduced to 0.51 of its no field value by 5000 oersteds.
[e] At 77°K the 200 difference reflection was reduced to 0.40 of its no field value by 5000 oersteds.
[f] At 77°K the 200 difference reflection was reduced to zero by 5000 oersteds.
[g] Additional structure is observed at positions inside the 200 reflection.
[h] At 77°K the ferromagnetic reflections are reduced to zero by 5000 oersteds. The additional structure is field independent.

termined by chemical analysis.[3] In the analysis a portion of each preparation was dissolved in a known excess of standard ferrous sulfate solution and the excess ferrous iron was titrated with standard potassium permanganate solution to give the oxidizing power of the manganese in the sample. A second portion was dissolved in nitric acid and hydrogen peroxide and the manganese oxidized to permanganate by sodium bismuthate. A known amount of ferrous sulfate was introduced into the solution and the excess determined by standard permanganate to give total manganese. From the results of the two titrations the ionic concentration was computed. The ionic analyses are shown in Column 3 of Table I and they are believed to be reliable to within about 5%.

Some consideration has been given also to the effect of quenching on the properties of the manganite preparations. In those cases where powder diffraction samples were to be used, the samples were quenched in air from the firing temperatures. When the briquet was to be used, it was necessary to cool the specimen slowly in order to avoid excessive fracturing. This was usually done by turning off the furnace and continuing to pass the firing atmosphere over the sample during cooling. Within the errors of analysis and observation, no significant differences in samples of comparable composition prepared by the two methods were observed.

[3] These analyses were carried out by Mr. A. D. Horton of the Analytical Chemistry Division.

3. APPARATUS

In the early part of this study the magnetic neutron scattering measurements were made with the conventional neutron spectrometer fitted with a low temperature cryostat which allowed data to be taken down to liquid helium temperatures, 4.2°K.

For a more complete analysis of the magnetic scattering properties of these compounds it was, however, found necessary to obtain data with the sample maintained in a magnetic field at low temperatures. The incorporation of magnetic fields at low temperatures required the building of a new spectrometer which is shown here in Fig. 1. The magnet and cryostat assembly sits on the crystal table of a low sturdily built spectrometer. The counter (for which only a part of the support arm is shown) rotates around the center of the crystal table and the counter and crystal table are geared together in the usual 2 to 1 ratio. The flat character of this spectrometer was dictated by the small distance from the floor to the center of the neutron beam from the pile.

The lower part of the cryostat vacuum chamber consists of a brass box sealed around the pole pieces with *O*-rings. The front and rear faces are *O*-ring sealed aluminum cover plates with thin walled cylindrical sections for entrance and exit of the neutron beam from the crystal monochromator. The front cover plate is removed for sample mounting and alignment. The inner cryostat and sample assembly are held in accurate

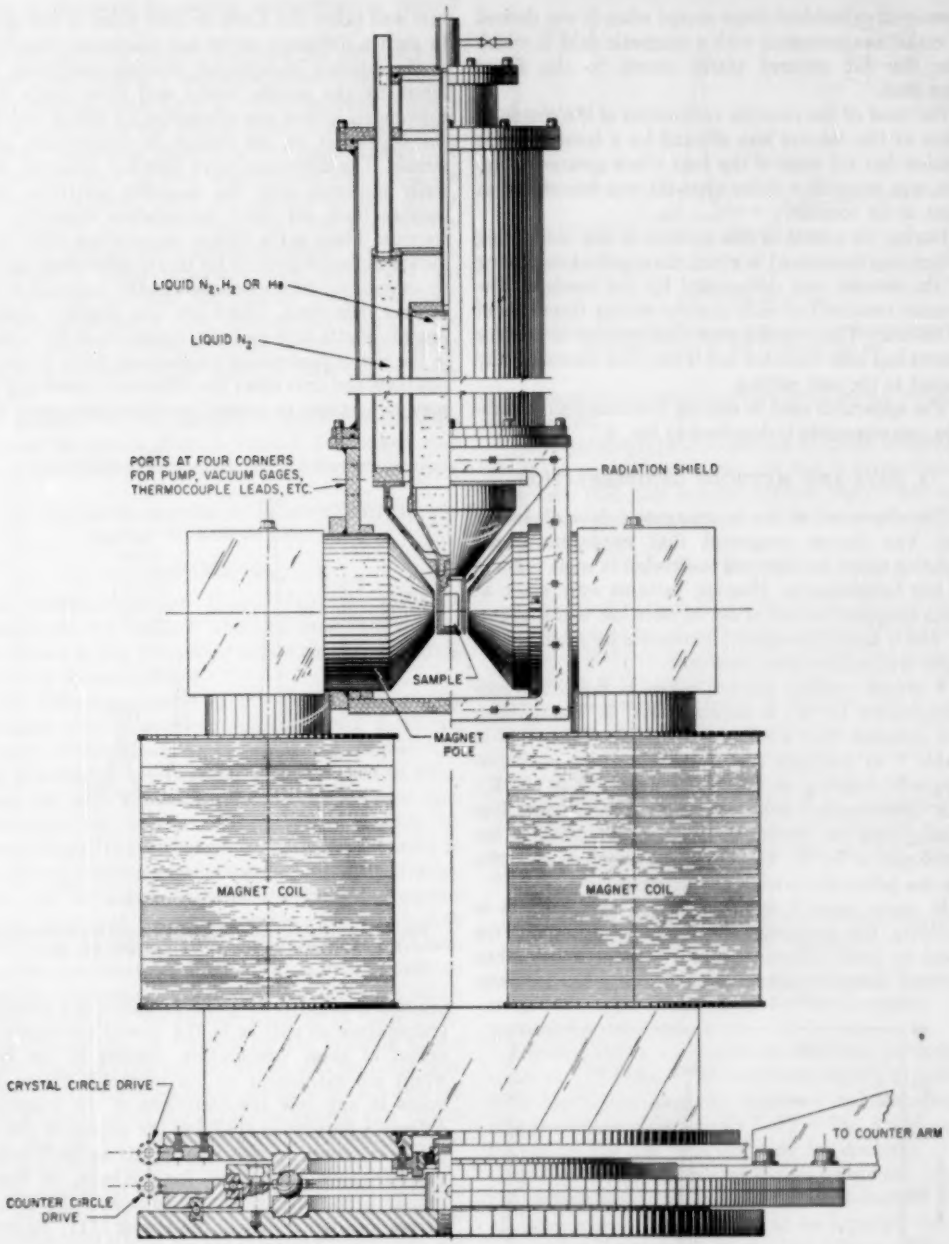


FIG. 1. Neutron spectrometer with cryostat and magnet.

alignment with the beam and the pole pieces by three nylon cords held taut by banjo type string tighteners anchored to the brass box.

The sample is cooled as previously by contact with the central dewar which for the lowest temperatures contained liquid helium. A thin-wall aluminum cylinder attached to this inner dewar and surrounding the sample constituted the inner radiation shield. This was in turn surrounded by a radiation shield in contact with the liquid nitrogen chamber. The samples were fre-

quently of cylindrical shape except when it was desired to make measurements with a magnetic field in which case the flat sintered plates shown in the figure were used.

For most of the runs the collimation of the scattered beam at the counter was effected by a boron carbide window but for some of the runs where greater resolution was required a Soller type slit was introduced in front of the counter.

During the course of this problem a new monitoring system was introduced in which the angular programing of the counter was determined by the monitor. The counter remained at each angular setting (intervals of 15 minutes of arc) until a prescribed number of monitor counts had been recorded and it was then automatically moved to the next setting.

The apparatus used in making ferromagnetic saturation measurements is described in Sec. 9.

4. DATA AND METHODS OF OBSERVATION

The character of the ferromagnetic data of Jonker and Van Santen suggested that antiferromagnetic ordering might be expected to develop in some samples at low temperatures. Neutron patterns were taken at room temperature and at one or more low temperatures at which antiferromagnetic structures as exhibited by super lattice lines were observed.

A sample neutron pattern taken at liquid nitrogen temperature (77°K) is shown in Fig. 2. This pattern was obtained with a sample of $LaMnO_3$ (identified in Table I as $LaMnO_3$ No. 1) which shows antiferromagnetic ordering at low temperatures ($T_c = 140$°K). The cross-hatched reflections are superstructure lines arising from the antiferromagnetic ordering which has developed at 77°K. The other reflections in this case are due primarily to nuclear scattering.

In many cases it has been found advantageous in studying the magnetic scattering to determine the point by point difference curves between a run taken at room temperature and one taken at a low temperature well below the Curie or Néel point of the sample. In such a difference curve one eliminates effects such as the external background, multiple scattering, scattering by the sample holder and other parts of the equipment, in fact one eliminates all effects which are not dependent on the change in temperature of the sample. The difference curve does not, however, necessarily represent only the magnetic scattering effects because there are other temperature dependent phenomena. There is the Debye temperature effect on the nuclear reflections which for strong reflections can show up as residual difference peaks at the positions of these nuclear reflections. There are also possible crystallographic effects such as temperature dependent changes in the lattice parameters which cause shifts in the peak positions and thus effect the difference curves and there may be changes in atomic position parameters which produce changes in the peak intensities as a function of temperature as well as in the crystal symmetry. The extent of these temperature changes in the crystal which are extraneous to the magnetic effects determines in any case the usefulness of the temperature difference pattern in studying the details of the magnetic scattering. In representing the antiferromagnetic superstructure reflections, for example, it has not usually been found profitable to extend the difference curves into the region of the strong (111) nuclear reflection because effects of nonmagnetic origin in some cases become prominent in this region. The method of representing the data will thus depend on the properties of the particular sample being studied. The ferromagnetic effects can also be studied by the alternative approach of utilizing a magnetic field to turn off the ferromagnetic scattering. The application of this tech-

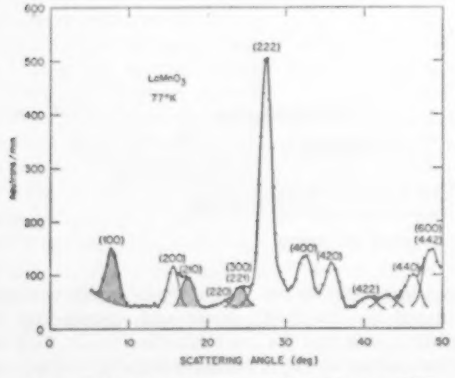


FIG. 2. Diffraction pattern for $LaMnO_3$ No. 1 at 77°K.

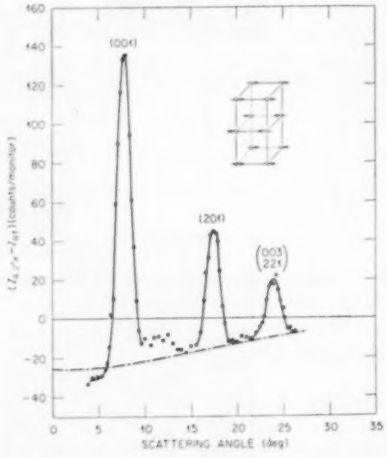


FIG. 3. Temperature difference pattern and antiferromagnetic structure (Type *A*) for $LaMnO_3$ No. 12.

nique is discussed in various connections in Secs. 7 and 8.

In the plot of a pattern involving the difference between a run taken at a temperature below the Curie or Néel point and a run taken at room temperature, where this is well above the critical point, the diffuse paramagnetic scattering features of the room temperature pattern will show up as a negative contribution in the difference pattern. There are features of this diffuse paramagnetic type scattering which are undoubtedly important to a complete understanding of the magnetic properties of these perovskite-type compounds but since they do not bear too directly on the magnetic structural aspects of these compounds, which is the primary subject of this paper, the details of the diffuse scattering features will be reserved for a future publication.

A summary of the samples studied and the data obtained for each is given in Table I. Discussions relative to these data are taken up in the following sections.

5. ANTIFERROMAGNETIC STRUCTURES OF $La^{+3}Mn^{+3}O_3$ AND $Ca^{+2}Mn^{+4}O_3$

$La^{+3}Mn^{+3}O_3$

A difference pattern ($I_{4.2°K}-I_{RT}$) for $LaMnO_3$ (Sample No. 12, Table I) obtained from data such as that shown in Fig. 2 together with a room temperature pattern is shown in Fig. 3.

The diffraction peaks[4] of this pattern are readily identified with an antiferromagnetic lattice consisting of layers of manganese ions coupled ferromagnetically (via intervening oxygen ions) in a given set of (001) planes but with alternate planes having opposite spin orientation, i.e., with antiferromagnetic coupling between planes. This structure is shown by the inset in Fig. 3, and as the *A*-type cell in Fig. 18. This arrangement calls for a unit magnetic cell doubled along one axis of the ideal elementary cubic cell. The indices identifying the reflections in the figure are nevertheless based on a cell doubled along all three axes in order to maintain uniformity in indexing later patterns which require increasing the cell in two or in all three directions. In addition to determining the ordering of the atomic moments one can determine also from the peak intensities for this cell whether the moments are directed approximately along the antiferromagnetic axis (taken as *Z* axis see Fig. 18) or in a plane perpendicular to this axis (*X*-*Y* plane). Since for the former case one would observe no intensity in the (001) reflection (see last column of Fig. 18 for predicted intensities) whereas experimentally this reflection is found to be strong, the moments must lie in or nearly in the *X*-*Y* plane. Although one might expect the moments to lie along one of the *X* or *Y* cube edges, this point cannot be ascertained in this case from powdered crystal measurements alone.

[4] The small peak near $2\theta=12°$ can be due to nuclear effects arising, for example, from shifts in oxygen parameters. This peak was not observed in patterns of $LaMnO_3$ of slightly higher Mn^{+4} content.

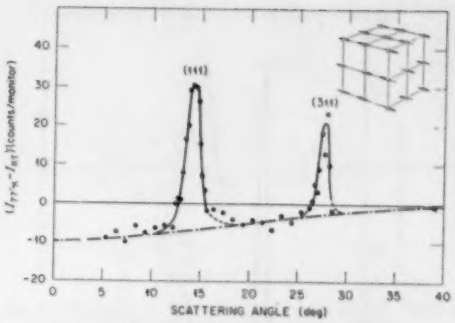


FIG. 4. Temperature difference pattern and antiferromagnetic structure (Type *G*) for $CaMnO_3$ No. 26.

The crystallographic properties of these compounds are taken up in detail in Sec. 11, but it is pertinent to note here that this sample departs from cubic symmetry in a manner consistent with the observed magnetic structure.

$Ca^{+2}Mn^{+4}O_3$

A difference pattern for $CaMnO_3$ No. 26 is shown in Fig. 4. It consists of two reflections which index as (111) and (311) on the basis of a cell doubled along the three edges of the simple chemical cell. This pattern is readily interpretable in terms of the structure given in the inset of Fig. 4, and identified as the *G*-type in Fig. 18 where the expected magnetic intensities are shown in the last column. Each Mn^{+4} ion is surrounded by six Mn^{+4} neighbors whose spins are antiparallel to the given ion. The structure can be thought of as consisting of two interpenetrating face centered lattices with opposite spin orientation.

6. OTHER ANTIFERROMAGNETIC STRUCTURES—EVIDENCE FOR ION ORDERING

In the composition range in which the Mn^{+4} content is greater than 50%, other antiferromagnetic type patterns are observed.

Figure 5 shows a temperature difference pattern for a sample [(0.20 La−0.80 Ca)MnO_3 No.74] with about 80% Mn^{+4}. The magnetic structure corresponding to this pattern is shown in the inset of Fig. 5 and as the *C*-type in Fig. 18. This cell calls for doubling of the simple cell in two directions but as before the lines in the pattern have been indexed for a cell doubled along all three axes. The intensities to be expected for this magnetic cell are shown in the last column of Fig. 18 for the two cases of spins along *Z* or in the *X*-*Y* plane. The intensities in the observed pattern correspond closely to case of spin parallel to the *Z*-axis.

In the *C*-type cell all atoms have two ferro- and four antiferromagnetic nearest neighbors whereas the reverse is true for the *A*-type cell.

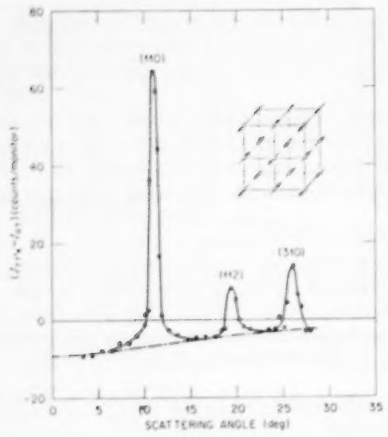


FIG. 5. Temperature difference pattern and antiferromagnetic structure (Type C) for (0.2 La−0.8 Ca)MnO_3 No. 74.

Up to this point antiferromagnetic patterns have been discussed for which the magnetic cells required doubling of the simple chemical cell along one, two, and three directions. For completeness in this respect one can add the ferromagnetic cell (discussed in the next section) for which the simple chemical cell and the magnetic cell are identical. These magnetic cells correspond to cases *A*, *B*, *C*, and *G* of Fig. 18. In each of these structures the ferro- and antiferromagnetic neighbor relations are the same for all the atoms in a given cell but the type of coupling and the relative number of ferro- and antiferromagnetic bonds varies with cell type.

We will consider next the pattern shown in Fig. 6 for a $CaMnO_3$ sample No. 6 having an Mn^{+4} content of about 70%. This pattern was obtained in the course of trying to prepare a pure $Ca^{+2}Mn^{+4}O_3$ sample. One notices here superstructure lines at the (100), (110), and (111) positions the simultaneous occurrence of which cannot be accounted for on the basis of any one of the previously encountered structures and also the intensities of these reflections do not correspond to the sum of the intensities from any incoherent mixtures of these structures. In view of this we were led to consider the other possible antiferromagnetic cells (*D*, *E* and *F*, Fig. 18) which require only doubling of the simple cell dimensions. These cells seem, *a priori*, no less acceptable than the previously observed cell types. Their magnetic neighbor and bond relations and the intensities in their diffraction patterns are also shown in Fig. 18.

The intensities in the pattern of Fig. 6 can be quite well accounted for as a mixture of magnetic phases of types *D* and *F* but at this stage we cannot be sure of the uniqueness of this interpretation. Further evidence for the incoherent mixing of magnetic phases will be seen to exist in connection with the composition range from greater than 0 to less than 50% Mn^{+4} content where both antiferromagnetic and ferromagnetic contributions to the pattern are simultaneously observed.

Further study of samples in this composition region ($x>0.5$) has, however, shown that some patterns cannot be accounted for by any of the cell types *A* to *G* or as incoherent mixtures of these magnetic structures. The first evidence for a more complicated magnetic structure came from a pattern which showed a broad peak at small angles when observed with the normal angular resolution. The difference pattern for sample (0.5 La−0.5 Ca)MnO_3 No. 66 taken with Soller slits to give better resolution is shown in Fig. 7.

The smallest unit cell upon which it is possible to index the observed magnetic reflections of this pattern is one which has $a=b=\sqrt{2}c$ where c is twice the Mn-Mn separation. The new a and b axes are thus face diagonals of the double cell (see inset in Fig. 18). The number of observed reflections which represents only a small fraction of those possible for this cell fall into two groups. In the one, single reflections (excluding permutations of indices) are observed with l odd, (101), (121), (301), etc., and in the other, contributions to a given peak can arise from superimposed even and odd l index reflections, (011) and (200), (311), and (202), etc. No single reflections with even l index are found nor are any reflections observed for planes of the form (001). Attempts to interpret the pattern with nonvanishing (200) and (202) reflections were not successful and hence the reflections in the latter group were ascribed to the (111) and (311) reflections only. On this basis all the observed reflections have l odd and this calls for a cell with all ions antiferromagnetically coupled in the c direction. The above considerations and a consideration of other absent reflections together with the fact that planes which contribute intensity cannot have a vanishing net spin lead to the arrangement of spins shown in the inset of Fig. 18.

It is interesting to observe that this spin arrangement can be developed by the coherent stacking of octants of the *C* and *E* type structures as shown by the octant

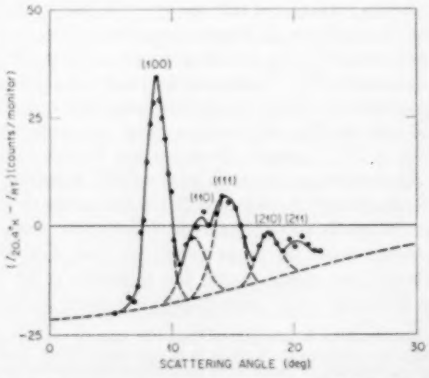


FIG. 6. Temperature difference pattern for $CaMnO_3$ No. 6.

labels in the structure diagram of Fig. 18. This stacking feature has led to the designation of this structure as the *C-E* type.

Although no other structure of this cell size has been observed, there are other possibilities two of which are represented by the *D-E* and *D-F* structures shown in the lower part of the figure. In the *C-E* type structure, which has been experimentally observed, the same simple ferro- and antiferromagnetic neighbor relations are present as in the case of the *C*-type alone, whereas the *D-E* and *D-F* cases are much more complicated in this respect.

So far the discussion of antiferromagnetic structures has been concerned only with the arrangement of the spins on the lattice sites and where possible with the orientation of the spin moments relative to the crystalline axes. In the case of the mixed oxides the magnetic structures may be associated with an ordering of the Mn^{+3} and Mn^{+4} ions on the lattice sites. In the *A*- and *G*-type structures this point did not enter. For the *C*-type structure the intensities are rather insensitive to ion ordering and no direct information on this point could be obtained from the data.

This is not true, however, for the *C-E* structure for which the predicted intensities are for some cases quite sensitive to the assumed distribution of the Mn^{+3} and Mn^{+4} ions on the lattice sites. Although this subject of ion ordering will be discussed later for all the mixed compounds in its relation to the nature of the magnetic coupling between ions, the experimental evidence for ion ordering will be presented here.

Before considering the effects of ion ordering it will be well to establish independently if possible the orientation of the spins relative to the crystalline axes. Calculations relative to both of these points are brought together in Table II. The first column gives the indices of the observed reflections based on the *C-E* type structure. The second column lists the values of q^2jF^2 deduced from the measured intensities in the pattern and succeeding columns list calculated values of this quantity. The first group of calculated values refers to a structure in which the Mn^{+3} and Mn^{+4} ions in equal proportions are distributed at random on the various lattice sites. Columns *a*, *b*, and *c* refer to the axis to which the spin vector has been assumed to be parallel.

TABLE II. Comparison of observed and calculated jF^2q^2 values for *C-E* type structure. Numbers in parentheses refer to percent $Mn^{+3}-Mn^{+4}$. *a*, *b*, *c* refer to axis along which spin is oriented.

	q^2jF^2 (10^{-24} cm^2/atom)						
		Calculated					
		Disordered (50–50)			Ordered (50–50)	(42–58)	(44–56)
hkl	Observed	*a*	*b*	*c*	*a*	*a*	*a*
101	0.45	0.49	0.74	0.25	0.64	0.33	0.40
111	0.75	1.06	1.06	0.71	0.78	0.78	0.78
121	0.82	1.05	0.45	0.75	1.37	0.71	0.86
301	0.1	0.09	0.49	0.41	0.12	0.05	0.07
311	0.85	1.11	1.11	1.59	0.82	0.82	0.82

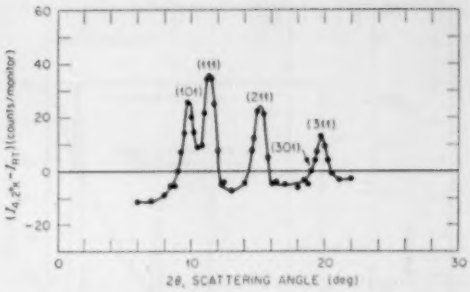


FIG. 7. Temperature difference pattern taken with Soller slits for (0.5 La−0.5 Ca)MnO_3 No. 66. *CE*-Type antiferromagnetic structure.

Although the agreement with the data is in no case very good, the low value predicted for the (301) reflection in the first column of this group suggests that the spins are oriented along the *a* axis. The fact that the calculations were made for a (50-50) ion composition whereas the sample analysis gave (42-58) has no effect on the relative values.

For the calculations involving ordering of the ions on lattice sites, two types of ordering have been considered.

One type results from simple phenomenological arguments about the types of indirect coupling between the magnetic ions in this system which are consistent with most of the data. The rules thus established which will be discussed more completely in Sec. 12 call for the indirect coupling in the region ($x>0.5$) to be ferromagnetic between unlike pairs and antiferromagnetic between like pairs of ions. The ion ordering obtained on the basis of these criteria puts Mn^{+3} and Mn^{+4} ions in positions for which the major intensity contributions are the same as for a disordered lattice. The ordering would lead to other nonvanishing reflections than those corresponding to the few observed lines in the pattern, but they would be too weak to have been observed.

The other type of ordering was suggested from a study of the pattern and it is of interest to observe that it fits well into a coupling scheme suggested by Goodenough (see following paper and Sec. 12 for further discussion on this point). This type of ion ordering is the one illustrated in the drawing of the *C-E* structure in Fig. 18.

For a pure structure of this type, the (101), (121), (301), etc., reflections would result from scattering by one kind of ion only and the (111), (311), etc., reflections would result from the other kind only. Unfortunately, the sample from which the pattern under consideration was obtained did not have a (50-50) ion composition, the results of the chemical analysis being (42-58). This fact makes it impossible to definitely verify this proposed type of ion ordering since additional assumptions must be made about the disposition of the excess Mn^{+4} ions.

Calculations relative to this type of ordering are

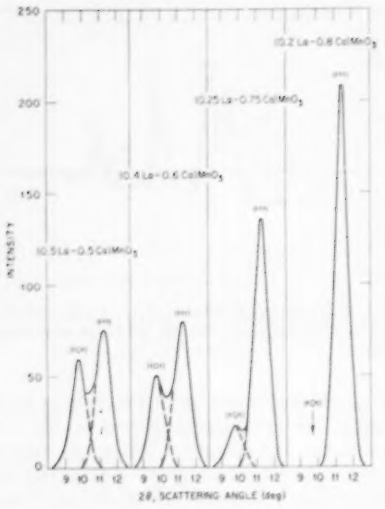


FIG. 8. First two antiferromagnetic reflections for four sample compositions showing the growth of a *C*-Type pattern (right) from the *CE*-Type (left).

given in the last three columns of Table II. The first column was calculated for the ideal composition. For the other two columns correction for composition was made on the assumption that the excess Mn^{+4} ions go at random but with opposite spin into sites occupied by Mn^{+3} ions in the ideal structure. The last column which agrees well with the experimental data was calculated for a slightly different composition than that given by the analysis but within the errors of the analysis.

The assumption that excess Mn^{+4} ions go into the Mn^{+3} sites with reversed spin is not completely arbitrary. If one examines the *C-E* type arrangement, one observes that one can obtain a *C* type structure by reversing half the spins of ions on the Mn^{+3} sites (as shown in Fig. 18) provided specific spins are so reversed. If this spin reversal is a result of the addition of Mn^{+4}, then one would expect that the *C* type structure should set in at a 75% Mn^{+4} concentration. Such a structure is in fact observed at 80% Mn^{+4}. It would seem reasonable to assume that in the early phases of this transformation the excess Mn^{+4} would go into the Mn^{+3} sites at random. A series of observations which suggested that this model for the transformation may be correct is shown in Fig. 8, where the first pair of reflections (101) and (111) are represented for various ion concentrations over the range 50%–80% Mn^{+4}. The intermediate compound (75 La−25 Ca)MnO_3 No. 18 is interesting in that it shows a much increased (111) reflection and a correspondingly decreased (101) reflection, and this indicates the approach to the *C*-type structure.

Before leaving the discussion of the *C-E* structure there is one point which should be mentioned. The discussion above relative to the deduction of the spin arrangement in this structure has assumed a single magnetic phase. Under the conditions that the structure is in fact defined by the unit cell described above and that there is antiferromagnetic coupling between adjacent layers, one can by exploring all of a reasonable number of possible spin arrangements conclude that the spin arrangement mentioned is unique. However, one may also put measurable intensities into and only into the positions observed in the pattern of Fig. 7 by an incoherent mixing of two phases: (1) the *C*-type which will reflect into the (111), (311), etc., position and (2) a cell of the same dimensions as the *C-E* type but built up of oriented octants of *E* type only (not illustrated) which may be called for the purpose of this discussion the *E-E* type and which puts intensity into the (101), (301), etc., positions. If one makes reasonable assumptions about the moment direction in the *E-E* arrangement and attempts to fit the observed intensities as an incoherent mixture of phases, one is led to inconsistencies which suggest the improbability of the incoherent arrangement. Further evidence from x-ray line splitting at low temperature which tends to support the *C-E* type arrangement is discussed in Sec. 12.

7. FERROMAGNETIC PATTERNS

One would expect from the saturation data of Fig. 12 and it is found from the neutron diffraction patterns that samples having a Mn^{+4} content in the neighborhood of 30% exhibit essentially pure ferromagnetic scattering with approximately the full predicted ferromagnetic moment per atom. Although the neutron patterns for sample compositions above and below this region also show ferromagnetic scattering, the situation becomes more complicated and these cases will be discussed in the following section.

Using the technique employed in the analysis of the antiferromagnetic patterns of plotting the difference between the low temperature and the room temperature data we obtained for sample (0.65 La−0.35 Ca)MnO_3 No. 70 the results shown in the lower half of Fig. 9. This pattern represents all the temperature dependent effects which occur between the two temperatures at which the patterns were taken, the greater part of which arises from magnetic scattering. The coherent peaks in this difference pattern fall at the positions of the nuclear reflections (the room temperature intensities of which are given above the pattern) as is to be expected for ferromagnetic scattering and the moments calculated from the two first peaks agree closely with the expected value. The third peak, however, is much too large to be due to magnetic scattering alone. There is, of course, the temperature effect for the strong nuclear peak at this position but on the basis of the *D-W* formula this would only correspond to that shown by the cross hatched portion for this and the succeeding peaks. One is thus led to conclude that minor changes

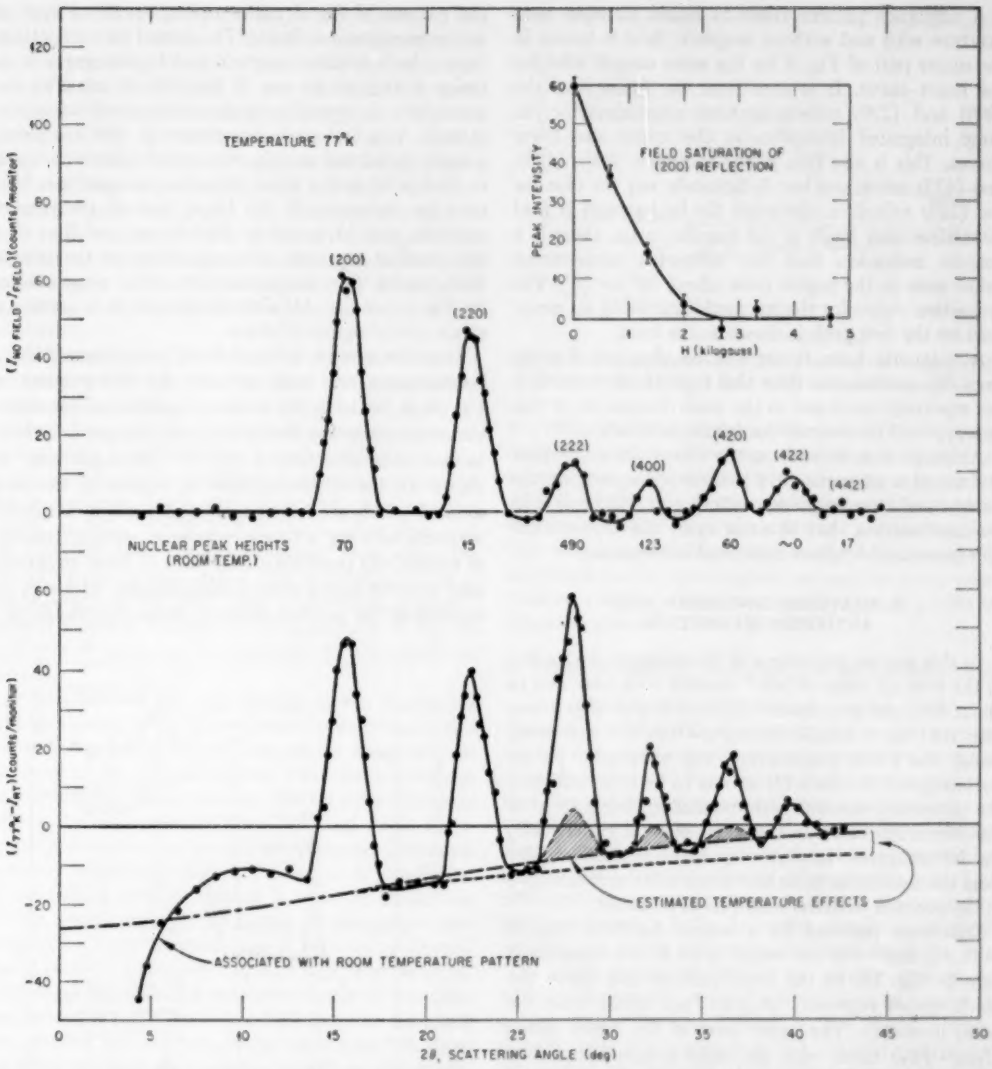


FIG. 9. Ferromagnetic patterns for [0.65 La−0.35 Ca]MnO_3 No. 70. Upper curve gives magnetic field difference pattern for maximum field in inset curve. Lower curve gives comparative data from temperature difference patterns.

with temperature of the atomic parameters are probably responsible for the anomalously large value of this peak.

The dashed curve represents the expected room temperature paramagnetic scattering. The anomalous behavior of the measured curve in the region from 4° to about 14° is undoubtedly due to a short range order effect in the room temperature pattern. Other patterns with a ferromagnetic component have shown a tendency to drop below the predicted curve at small angles and this effect has been identified with the room temperature data.

It must be concluded that the low temperature-room temperature difference pattern does not give an entirely reliable measure of the ferromagnetic scattering. In most cases the best approach to the study of the ferromagnetic scattering contribution is to use a magnetic field. Since the magnetic scattering goes to zero when the moments are all aligned parallel to the scattering vector the difference between measurements taken with zero field and with a field along the scattering vector of sufficient strength to align the domains gives a direct measure of the ferromagnetic scattering.

A difference pattern taken at liquid nitrogen temperature with and without magnetic field is shown in the upper part of Fig. 9 for the same sample used for the lower curve. It is seen from the figure that the (200) and (220) reflections have approximately the same integrated intensities in the upper and lower curves. This is also true of the corrected (400), (420), and (422) reflections but is definitely not the case for the (222) reflection. Although the background is field insensitive over much of the angular range, there is a definite indication that the difference curve drops below zero in the region from about 25° to 35°. The saturation curve for the magnetic scattering as measured on the first peak is shown in the inset.

A magnetic form factor can be obtained directly from this pattern but since this type of information is not especially pertinent to the main discussions of this paper, it will be reserved for future publication.

Although it appears from the above discussion that the use of a magnetic field is ideal for separating the nuclear and ferromagnetic scattering, it will be seen in the next section that in some cases the temperature difference method gives additional information.

8. MIXTURES OF FERRO- AND ANTIFERROMAGNETISM

In this section the nature of the magnetic scattering in the over-all range of Mn^{+4} content from near zero to about 50% will be considered. In the last section it was seen that for a sample having $x=0.35$ the scattering below the Curie temperature was essentially purely ferromagnetic in character as was to be expected since the saturation moment corresponded approximately to the theoretical value. In the region of $x<0.35$, however, the ferromagnetic saturation moment falls off rapidly from the maximum value and hence some modifications in the neutron patterns would be anticipated.

Difference patterns for a sample $LaMnO_3$ No. 73 -17.5% Mn^{+4} [8] in the region $x<0.35$ are reproduced here in Fig. 10. In the lower part of this figure the black circles represent $(I_{4.2°K}-I_{RT})$ taken with the usual geometry. The upper part of the figure shows $(I_{77°K}-I_{RT})$ taken with the same sample but with a Soller slit in front of the counter to give the improved resolution. The (200) and (220) reflections in these difference curves would be expected to arise essentially entirely from ferromagnetic scattering as illustrated by Fig. 9 for the purely ferromagnetic case. The cross hatched peaks (100) and (210), however, are at superstructure positions, in fact they correspond both in their positions and their relative intensities to those in the pattern of Fig. 3 corresponding to the A type of antiferromagnetic ordering. To account for this pattern having both antiferromagnetic and ferromagnetic scattering contributions one is first led to consider the possibility of crystallographic inhomogeneities in the sample. As a first crude experiment to test this point, a finely powdered sample was cooled below the Curie or Neél point and a small permanent magnet was held near the container. It was found that all the powder particles were attracted by the magnet and thus that any possible magnetic inhomogeneities in the sample were smaller than the particle size. Later x-ray studies by Yakel (see Sec. 11) showed the sample to possess a single crystallographic phase.

Since the sample is found to be crystallographically homogeneous one must account for the pattern of Fig. 10 on the basis (a) of some magnetic cell for which the moments in one direction do not compensate those in the other direction, a type of ferrimagnetism, or (b) of an incoherent mixture of regions or domains some of which are ferromagnetic and others antiferromagnetic with the A type structure; in addition there is, of course, the possibility that both of these situations may to some degree exist simultaneously. Although an analysis of the problem seems to favor hypothesis (b),

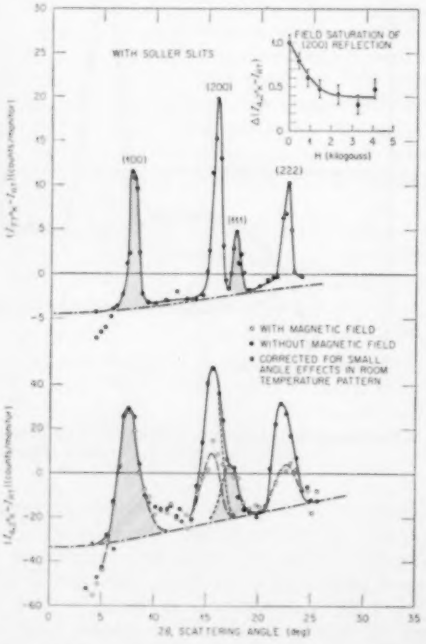


FIG. 10. Difference patterns for $LaMnO_3$ No. 73 with mixed ferro- and antiferromagnetic phases. Cross hatched lines represent antiferromagnetic contribution, others ferromagnetic. Upper curve taken with Soller slits, lower curve taken with usual resolution with and without magnetic field. Intensity *vs* field shown in inset.

[8] Although the ideal sample for this Mn^{+4} content should have the composition $(0.83\ La-0.17\ Ca)MnO_3$ it has been found that in this composition range the patterns are sensitive only to the Mn^{+4} content. For example, sample No. 23 shows essentially the same characteristics as sample No. 73. The latter sample is used here because it was made in the briquet form required for the measurements with a magnetic field.

at least as representing the major contributor, it was felt that a direct check of this point was important and it was primarily for this reason that the new spectrometer described in Sec. 2 was constructed. If ferromagnetism and antiferromagnetism appear in separate domains which are uncoupled or very weakly coupled, it would be expected that a magnetic field applied along the scattering vector would turn off only the ferromagnetic scattering peaks. If, however, the pattern were to be accounted for in terms of a homogeneously coupled system of some ferrimagnetic type the field would couple to the whole system and all the magnetic scattering could be turned off within the limits of field saturation.

The lower part of Fig. 10 shows the result of such a field study on this sample. The black circles, as we have mentioned, represent the difference curve ($I_{4.2°K}-I_{RT}$) with no field and the open circles represent the same pattern with a field of 4500 oersteds applied along the scattering vector. It is evident from these curves that the superlattice peaks are unaffected within experimental error by the field. It is also evident that with the available field the ferromagnetic type reflections are only partially obliterated. The saturation nature of the approach of the (200) peak height to its minimum value is shown as the inset in the upper part of Fig. 10. This type of curve was obtained only for a temperature of 77°K.

A discussion of the significance of the incomplete field sensitivity of the ferromagnetic type peaks in this and other patterns in the composition range will be discussed in Sec. 9 in connection with direct ferromagnetic saturation measurements. The fact that the magnetic measurements show the (200) and (220) ferromagnetic reflections to be field sensitive and the superstructure reflections (100) and (210) to be unaffected by the field is strong evidence in favor of accounting for the overall magnetic structure of samples in this composition range as consisting of an incoherent mixture of ferro- and antiferromagnetic regions or domains.

Although the data for only one sample in the composition range of Mn^{+4} less than 35% has been discussed here, several samples have been investigated to show the trend of the mixing in the range from the purely antiferromagnetic phase to the essentially purely ferromagnetic phase discussed in the last section. The other cases are tabulated in Table I and the nature of the overall results will be taken up in Sec. 13 in connection with the summary curve Fig. 19 of the ferro- and antiferromagnetic moments *versus* composition.

To complete the region involving ferromagnetic scattering there is shown in Fig. 11 a difference pattern for a sample (0.6 La−0.4 Ca)MnO_3 No. 4 with an Mn^{+4} content of 44%. This pattern shows primarily ferromagnetic scattering as represented by the main peaks but one sees also a small amount of antiferromagnetic contribution in the angular range 7 to 15 degrees and in the neighborhood of 20 degrees. Although this

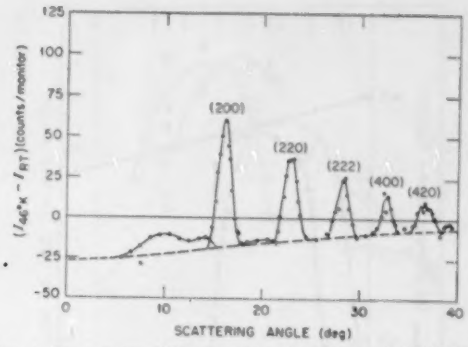


FIG. 11. Temperature difference pattern for (0.6 La−0.4 Ca)MnO_3 No. 4 for which there is some antiferromagnetic scattering.

scattering is too weak to permit identification with a definite magnetic structure, there is good reason to believe from comparisons with other patterns taken with the same and better resolution that this corresponds to the *C-E* type of antiferromagnetic structure. One thus has most likely also on the high side of the maximum in the ferromagnetic saturation *versus* composition curve a region of mixed ferro- and antiferromagnetic phases.

9. FERROMAGNETIC SATURATION MEASUREMENTS

The ferromagnetic properties of samples in the low range of Mn^{+4} content as determined from the neutron studies did not fit too well with the saturation moment data of Jonker and Van Santen and hence it was felt that direct measurements of the saturation moments on the samples used in this study would be valuable. Since a double cryostat surrounded at its lower end by an oil cooled solenoid capable of producing a field of 5000 oersteds constant over several inches was available in the laboratory,[6] the following simple method of making saturation magnetization measurements was used which tests indicated to be reliable at least to within a few percent. A small coil of 100 turns wound on a thin wall tube was located in the liquid helium cryostat near the center and with its axis parallel to the field of the solenoid. This tube projected through the top of the cryostat and another loosely fitting concentric tube was used to hold a sample approximately coaxial with the 100 turn coil. The powdered samples contained in cylindrical cells 2 in. long by ¼-in. diameter were held in the center of the coil until they had come to thermal equilibrium with the cooling liquid. The measurements were made by withdrawing the samples and reading the deflection of a ballistic galvanometer connected to the coil. The readings were standardized by comparison with the readings from a powdered Ni sample. For those samples for which these measurements were made, the results are tabulated in

[6] We are indebted to Dr. L. D. Roberts for the use of this equipment.

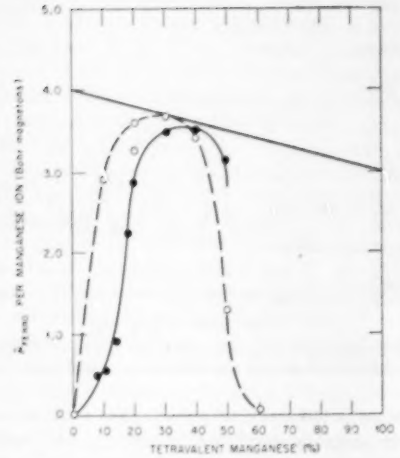


FIG. 12. Ferromagnetic saturation data vs composition. Open circles Jonker and Van Santen, solid circles present work.

the last column of Table I as the average moment per atom in Bohr magnetons. These data are shown also in Fig. 12 together with the corresponding data of Jonker and Van Santen. It is to be noted that our measurements are lower than those of Jonker and Van Santen in the region of low Mn^{+4} content and this point will come into evidence again later when a comparison is made with the corresponding neutron diffraction results.

An additional aspect of these measurements relates to the nature of the *B-H* curves. There would seem to be no question about temperature saturation since the measurements were made at liquid helium temperature. There is, however, a problem in some samples about field saturation. Those samples such as No. 70 which have a high moment are found to behave very normally reaching saturation at low fields. For the case of sample No. 65 with ~14% Mn^{+4} content (and for those of lower Mn^{+4} content) unusual *B-H* curve characteristics are observed as illustrated in Fig. 13. There seem here to be at least two types of magnetic behavior and it would appear also that saturation is not complete at 5000 oersteds. The peculiar character of this *B-H* curve has a bearing on the data to be discussed in Sec. 19 but the nature of the magnetic phenomenon involved has not been determined.

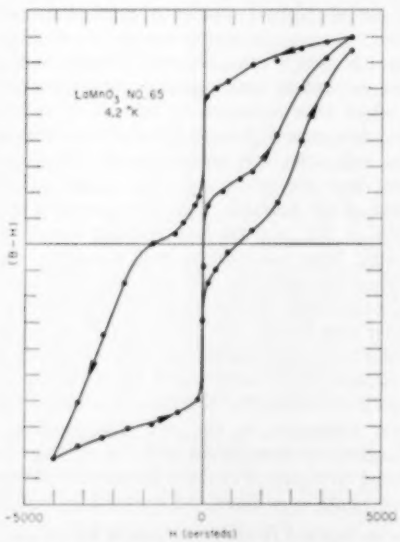


FIG. 13. *B-H* curve for $LaMnO_3$ No. 73.

10. THE CURIE OR NÉEL TEMPERATURE

Since the intensity of a coherent magnetic reflection in a neutron pattern depends upon the degree of temperature saturation of the magnetic lattice one can measure the temperature of transition to an ordered magnetic state by observing the temperature dependence of the intensity of a coherent magnetic reflection. Temperature-intensity curves were obtained for a number of the samples listed in Table I and a representative set of data is shown in Fig. 14. Intensity data were taken during a slow warmup from 77°K, and they were fitted by a Brillouin function, the intercept with the temperature axis of which was taken as the Néel temperature. The excess scattering at temperatures above the Neél point is probably due to short range magnetic ordering. This procedure was followed in cases for which 77°K was the lowest temperature at which patterns were obtained. In other cases, data at two or more fixed temperatures were extrapolated with a Brillouin function to give an estimate of the Néel or Curie point. In those instances where two phases are present, the transition temperature given in the table

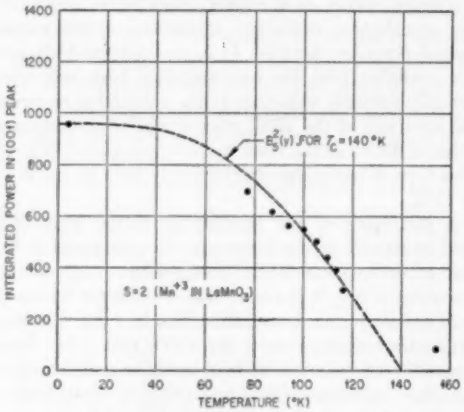


FIG. 14. Typical temperature saturation data with Brillouin curve.

refers to the dominant phase. In one case of a mixed phase sample ($LaMnO_3$ No. 3) the two transition temperatures appeared to be approximately equal. In another case ($LaMnO_3$ No. 73) there was an indication that the Neél temperature was some 30°K lower than the Curie temperature, but since the minor phase reflections are weak even at saturation, it is difficult to evaluate the minor phase transition points. The transition temperatures obtained by one or the other method outlined above are plotted in Fig. 15 in which are also shown the ferromagnetic Curie temperatures obtained by Jonker and Van Santen.

11. DETERMINATIONS OF APPARENT SYMMETRY AND LATTICE PARAMETERS

As a part of this investigation, x-ray diffraction data have been obtained by Yakel for all preparations at room temperature, and also at lower temperatures for a few selected samples. The room temperature results have already been described.[7] As Yakel has pointed out, the differences between the diffraction patterns obtained and those expected from an ideal perovskite structure consist of split reflections, and superlattice reflections which imply distortions of the lattice from cubic symmetry. Since, as shown by Megaw,[8] the number and relative intensities of the diffraction lines into which a given ideal reflection splits by distortion of the unit cube is characteristic of the symmetry of the distorted structure, it was possible to index the patterns and to determine the apparent symmetry of the preparations in most cases. In a few instances the splitting was so slight as to make the separate reflections almost indistinguishable, and in such cases only an average lattice parameter was measured.

The room temperature x-ray diffraction data are summarized in Fig. 16 in which are plotted the lattice parameters of our preparations against tetravalent manganese concentration. For tetravalent manganese concentrations up to about 25%, the splitting is characteristic of a monoclinic or orthorhombic symmetry, and this splitting decreases with increasing tetravalent manganese concentration. As represented in the figure, the apparent symmetry is taken as monoclinic, but since within experimental error the a_1 and a_3 axes of the monoclinic cell are equal, an orthorhombic cell with a_1 and a_3 axes equal to the diagonals of the rhombus formed by the monoclinic a_1 and a_3 axes may also be defined. Which of these correctly describes the symmetry of the unit cell cannot be decided from powder data alone. The splitting which is observed in the 0–25% concentration range seems to depend primarily upon the tetravalent manganese content of the preparation since mixed oxides and $LaMnO_3$ preparations with corresponding ion concentrations exhibit similar splittings. For the entire series of preparations except for the

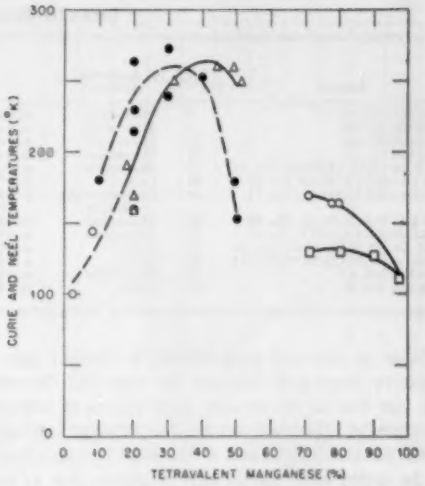


FIG. 15. Curie and Néel temperatures *vs* composition. Solid circles Jonker and Van Santen ferromagnetic saturation data, open triangles ferromagnetic neutron data, open circles and squares antiferromagnetic data, square referring to $CaMnO_3$ series.

$CaMnO_3$ samples with fairly high trivalent manganese concentration, the average lattice constant as defined by the approximate relation $\bar{a}=V^{\frac{1}{3}}$ (where V is the

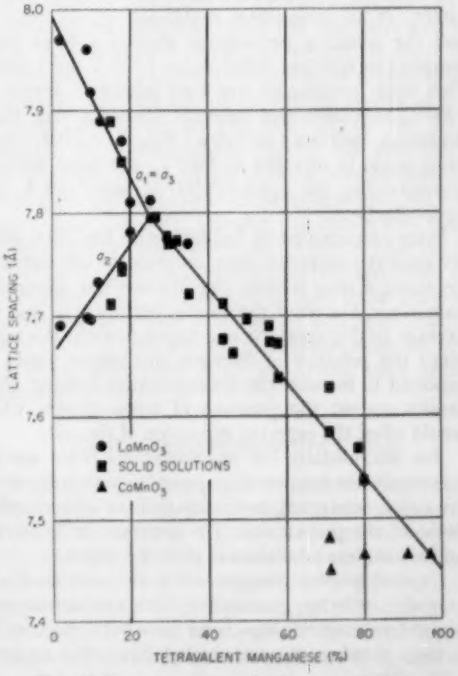


FIG. 16. Room temperature lattice constants from x-ray measurements. Yakel.

[7] H. L. Yakel, Acta Cryst. (to be published).
[8] H. D. Megaw, Proc. Phys. Soc. (London) 58, 133 (1946).

TABLE III. Summary of x-ray diffraction data.

Sample	R.T. Mn^{+4} %	Apparent symmetry	Lattice parameters	T (°K)	Low temp. Remarks
$LaMnO_3$ No. 12	2	Monoclinic	$a_1=a_3=7.973$ A, $a_2=7.693$ A, $\beta=92°\ 7'$	80	No further splitting
$LaMnO_3$ No. 34	9	Monoclinic	$a_1=a_3=7.960$ A, $a_2=7.698$ A, $\beta=91°\ 52'$	93	Very slight additional splitting
$LaMnO_3$ No. 65	12	Monoclinic	$a_1=a_3=7.891$ A, $a_2=7.737$ A, $\beta=91°\ 9'$	80	No further splitting
$LaMnO_3$ No. 3	20	Monoclinic	$a_1=a_3=7.809$ A, $a_2=7.782$ A, $\beta=90°\ 12'$	93	No further splitting
(0.85 La−0.15 Ca)MnO_3 No. 21	18	Monoclinic	$a_1=a_3=7.848$ A, $a_2=7.734$ A, $\beta=90°\ 45'$	83	Very slight additional splitting
(0.5 La−0.5 Ca)MnO_3 No. 16	49	Cubic	$a_1=7.697$ A	93	No splitting
(0.5 La−0.5 Ca)MnO_3 No. 75	46	Unmeasurable splitting	$\delta=7.662$ A	78	No further splitting
(0.5 La−0.5 Ca)MnO_3 No. 66	58	Monoclinic	$a_1=c_2=7.672$ A, $a_3=7.638$ A, $\beta=90°\ 12'$	78	Tetragonal $a_1=7.713$ A, $a_3=7.515$ A
(0.4 La−0.6 Ca)MnO_3 No. 8	71	Tetragonal	$a_1=7.628$ A, $a_3=7.584$ A	87	Tetragonal $a_1=7.668$ A, $a_3=7.508$ A
(0.25 La−0.75 Ca)MnO_3 No. 18	78	Cubic	$a_1=7.568$ A	81	Tetragonal $a_1=7.612$ A, $a_3=7.476$ A
(0.20 La−0.80 Ca)MnO_3 No. 74	80	Cubic	$a=7.539$ A	86	Tetragonal $a_1=7.466$ A, $a_3=7.636$ A
$CaMnO_3$ No. 6	71	Monoclinic	$a_1=a_3=7.481$ A, $a_2=7.449$ A, $\beta=91°\ 7'$	93	No further splitting
$CaMnO_3$ No. 26	91	Cubic	$a_1=7.465$ A	80	No splitting

volume of the unit cell) follows a straight line with negative slope and displays the expected decrease in cell size due to the smaller ionic radius of tetravalent manganese. This relation for the average spacing also holds for the compounds in the 0–25% region but it is to be noted that the a_2 axis increases, the a_1 and a_3 axes decrease over this region. It is therefore indicated that the $LaMnO_3$ preparations are characterized by cation defects rather than by interstitial anions but that the defects do not play a very important role with regard to the interatomic distances.

Some observations were also made which suggest that these defects adjust themselves to the atmosphere and temperature of firing and are formed reversibly. For instance a preparation of $LaMnO_3$ was fired at 1400°C in an atmosphere containing 1% O_2-99% N_2 and the resulting preparation showed a large room temperature splitting with $a_1=a_3=7.950$ A, $a_2=7.700$ A. This same preparation was then refired in oxygen at 1200°C and diffraction patterns showed a very much decreased splitting; $a_1=a_3=7.812$, $a_2=7.794$. After firing again in nitrogen at 1400°C, the large splitting corresponding to $a_1=a_3=7.960$ A, $a_2=7.698$ A was again observed.

Some preparations of $CaMnO_3$ with less than 100% tetravalent manganese were obtained in the course of preparing a pure sample. One of these was apparently monoclinic, the other three cubic but in every case the average lattice constant was approximately the same. Since the presence of trivalent manganese would be expected to increase the average lattice spacing, these results suggest the presence of anion defects which would offset the expected expansion of the cell.

For intermediate ion concentrations there are observed at room temperature apparent symmetries which are cubic, tetragonal, and monoclinic or orthorhombic. Some of the preparations are described in Table III, and a complete tabulation is given by Yakel.[7]

Crystallographic changes which are associated with magnetic ordering, particularly with the development of antiferromagnetic structures have been observed for a large number of compounds.[9] Since the magnetic ordering which develops in the manganites does so at low temperatures, a number of preparations have been examined by x-ray diffraction methods at temperatures slightly above 77°K. The results of this study are shown in Table III.[10] For specimens in the 0–50% concentration range, very little further splitting was observed. In the region of 55%–85% except for $CaMnO_3$ No. 6 an additional splitting of appreciable magnitude was detected. Of the four samples listed which show this effect, the first three exhibit the magnetic structure which we have designated as the *C-E* type, the fourth one belongs to the *C* type. In each case the splitting pattern suggests a tetragonal cell but in the first three instances the tetragonal axis decreases whereas for the *C* type magnetic structure the tetragonal axis increases over the room temperature value. The transition from the room temperature to the low temperature diffraction patterns for these substances is not sharp but extends over a large temperature range. A typical curve showing the variation of lattice parameter with temperature is shown in Fig. 17. One sees from this curve that the splitting increases with decreasing temperature and apparently levels off to a maximum at about 150°K. The Neél temperature for this sample as determined from the neutron data is approximately 165°K.

It is indicated that the crystallographic distortion which occurs, does so at a temperature considerably above the Neél point and it has been suggested by Goodenough[11] that the crystallographic distortion is due to an ordering of covalent bonds in such a way as to minimize the elastic distortions while permitting the formation of the optimum number of covalent bonds. In pursuing these arguments Goodenough has been led to infer from the c/a ratios (<1 for *C-E*; >1 for *C*) a system of ionic ordering which is in agreement with the neutron diffraction observations and thus lends support to our assignment of sample 18 to the *C-E* structure type.

The room temperature neutron patterns, even for the concentration range for which the apparent symmetry was found by x-ray measurements to be cubic, all

[9] See for instance the review article by Nagamiya, Yosida, and Kudo, Advances in Physics **4**, 2 (1955).

[10] H. L. Yakel (private communication).

[11] John B. Goodenough, following paper, Phys. Rev. **100**, 564 (1955).

showed evidences of distortion from the simple cubic perovskite structure. This evidence consists of (1) the lack of agreement between the neutron intensities observed and those calculated on the basis of the simple cubic perovskite structure and (2) in some cases weak but measurable lines at positions which require a doubling of the simple cell. Such effects have been noted also by Roth in a preparation of (La, Ba)MnO_3.[12]

The observed powder intensities can be better correlated if the oxygen ions are displaced from the special positions of the ideal perovskite cell, for instance, if the oxygen tetrahedra are rigidly rotated in the manner proposed by Naray-Szabo in the $CaTiO_3$ structure.[13] Other conceivable distortions of the oxygen ion system all tend to improve the intensity agreement. Therefore, powder data alone do not seem sufficient for a definite decision as to the structure distortions and a more complete study must be carried out with single crystals.

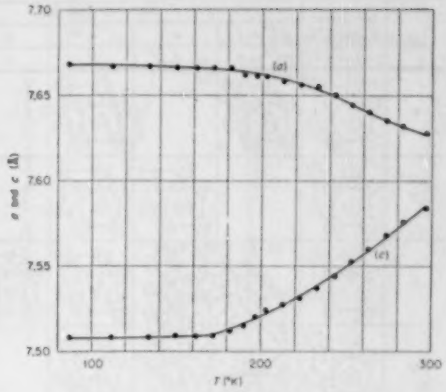


FIG. 17. Temperature dependence of lattice constants for a typical *CE*-Type sample. Lattice constants are twice those of the simple cubic cell.

12. SUGGESTED DISTRIBUTION AND COUPLING OF Mn^{+3} AND Mn^{+4} IONS IN VARIOUS MAGNETIC STRUCTURES

The correlation of the observed magnetic structures of this series of compounds with Mn^{+3} and Mn^{+4} ion concentration should give information about the type and strength of coupling between these ions. From the over-all structural aspects of the system one is led to the conclusion that the exchange coupling in general is (a) ferromagnetic between Mn^{+3} and Mn^{+4} ions (b) antiferromagnetic between Mn^{+4} ions and (c) either ferromagnetic or antiferromagnetic between Mn^{+3} ions, the type being related to the separation between ions.

It would also seem reasonable to suppose that long range ordering of the magnetic moments would be accompanied by some degree of ordering of Mn^{+3} and Mn^{+4} ions on correspondingly appropriate lattice sites. This ordering requires only that an electron be free to shift readily from one ion to another.

On the basis of these hypotheses a system of ordered magnetic structures can be developed as shown by the single octant entries in Columns 3 to 7 of Fig. 18. Pertinent points relative to these proposed structures will be considered in the order in which they appear. For convenience they will sometimes be labeled with the letter in the first column and the Mn^{+3}, Mn^{+4} content listed at the top of Columns 3 to 7. It will be seen in the following that one or more of the set of preliminary conclusions above must be relaxed for certain composition ranges.

Pure Mn^{+3} Phase ($x=0$)

In the ideal $LaMnO_3$ phase all manganese ions are trivalent and one must look to a crystallographic or bond character origin for the fact that there is ferromagnetic coupling along two axes and antiferromagnetic coupling along the third axis. There is indeed a striking correlation between the magnetic structure and the nearly tetragonal character of the crystal structure for $x=0$, the lattice parameters of which are shown in Fig. 16 for room temperature and which are not much altered for temperatures below the Neél point. The magnetic and crystal structure properties of this phase are both well accounted for by Goodenough's theory of bonding for these compounds. He associates with the Mn^{+3} ions four-lobed planar hybrid orbitals which can thus overlap the p orbitals of only four of the six oxygen neighbors. His proposal of an ordered system of such orbitals as represented schematically by the arrow heads in the *A*-8-0 entry in Fig. 18 will thus lead to the observed crystal symmetry if, as is to be expected, overlap on both sides of an intervening oxygen ion leads to a shorter Mn-O-Mn distance. His theory also makes overlap on both sides of an intervening oxygen ion the condition for antiferromagnetic coupling and hence this will correspond to the unique a_3 axis. From the x-ray measurements it may thus also be concluded that the Mn^{+3}-O-Mn^{+3} coupling will be antiferromagnetic when the separation of two Mn^{+3} ions is not much greater than 3.8 A.

Ferromagnetic Region ($0<x<0.5$)

With a relatively strong ferromagnetic coupling for Mn^{+3}-O-Mn^{+4} one is led to predict a stable ferromagnetic cell *B*-6-2 for a 25% Mn^{+4} content as shown in Column 4, Fig. 18. The presence of two Mn^{+4} ions per unit cell would then be sufficient to stabilize the ferromagnetic phase. With increasing Mn^{+4} content one would expect, if the ferromagnetic Mn^{+3}-O-Mn^{+4} exchange is dominant, the development of a structure *B*-4-4 in which every Mn^{+3} ion has six Mn^{+4} nearest neighbors. The behavior of the moments in the mixed

[12] Walter L. Roth, Pittsburgh Diffraction Conference, November, 1954 (unpublished).
[13] S. Naray-Szabo, Naturwiss. 31, 202 (1943).

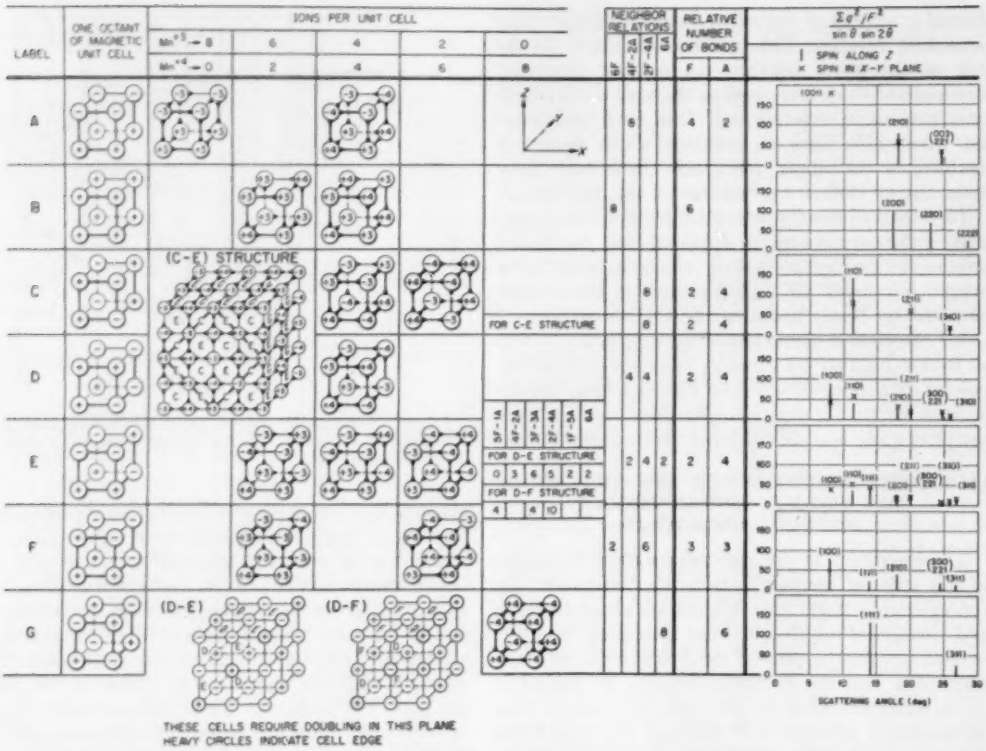


FIG. 18. Scheme of magnetic structures and related information. *A*, *B*, *C*, *G*, and (*C-E*) definitely observed and some evidence for *D* and *F*. Ion ordering schemes represent arrangements consistent with certain coupling criteria. Arrowheads are a schematic representation of Goodenough's semicovalent exchange coupling.

phase region ($0.1<x<0.25$) discussed in Sec. 13 indicates that there is a tendency for the Mn^{+4} ions to cluster and hence some tendency toward ordering of the type suggested by *B*-6-2.

Conversely, in the region ($0.25<x<0.35$) where the observed ferromagnetic moment is nearly equal to the total spin-only value; an ordered structure intermediate to *B*-6-2 and *B*-4-4 should give rise to a diffraction pattern exhibiting superlattice lines, the intensities of which depend upon the square of the difference of the magnetic scattering amplitude of Mn^{+3} and Mn^{+4}. These superlattice reflections will always be weak but in favorable cases should be just observable. No such superlattice lines were found and this suggests that the system is not completely ordered. A disordered arrangement of Mn^{+3} and Mn^{+4} ions over the lattice sites is consistent with the resistance minimum[14] observed near $x=0.3$ and with the fact that for these compositions the structure is nearly cubic.[11]

The sharp decline of the ferromagnetic moment as $x\to0.5$ suggests that there is a strong tendency for antiferromagnetic coupling beyond this region. This trend can be reasonably accounted for on the bases (a) that beyond $x=0.5$ Mn^{+4}-O-Mn^{+4} antiferromagnetic bonds develop (b) the lattice spacings for $x>0.5$ are less than or equal to the spacing in the *A* type structure which is associated with a Mn^{+3}-O-Mn^{+3} antiferromagnetic bond, and (c) the antiferromagnetic *C-E* type structure which satisfies very well the over-all coupling requirements for $x=0.5$ is known to develop at nearly this composition.

[14] J. H. Van Santen and G. H. Jonker, Physica **16**, 599 (1950).

Antiferromagnetic Region $x>0.5$

The various possible ordered arrangements of manganese ions in antiferromagnetic structures exhibited in Fig. 18 were postulated for the most part on the basis of the exchanges (a), (b), and (c) suggested in the first part of this section.

The structure type *A*-4-4 would be consistent with these exchange criteria if the Mn^{+3}-O-Mn^{+4} ferromagnetic exchange remained strong at this concentration. The fact that this structure is not observed for the $x=0.5$ composition may be due to other arrangements with fewer ferromagnetic bonds being more stable. For example the *C*-4-4 structure would also satisfy the requirements of the above exchange assumptions but

this also was not observed. The *C*-*E* structure however, is a satisfactory and interesting occupant of the 50% Mn^{+4} region. On the basis of the previous assumptions of the coupling between Mn^{+3} and Mn^{+4} ions the ordering would be a combination of the elements *C*-4-4, *E*-6-2, and *E*-2-6. It is this type of ordering which gives the intensities shown in the second column of Table II, which we have seen can be improved by the alternative ordering scheme, shown as the *C*-*E* arrangement in the inset of Fig. 18. This arrangement which calls for both ferro- and antiferromagnetic coupling in the Mn^{+4}-O-Mn^{+3} bond is in perfect accord with the semicovalent coupling criteria of Goodenough. The arrowheads in the diagram of this structure have been incorporated to bring out in a simple way some of the elements of the theory. Each arrow represents essentially the overlap of a hydridized manganese ion orbital with a neighboring oxygen p orbital. According to the theory overlap on both sides of the intervening oxygen ion gives antiferromagnetic coupling whereas overlap on only one side leads to ferromagnetic coupling. In the figure which is meant only to be suggestive of the ideas involved, Mn^{+4} ions are represented as having octohedral orbitals and Mn^{+3} ions as having square coplanar orbitals, wherever this is consistent with the coupling criteria. The details are given in the following paper.

With increasing Mn^{+4} content for $x>0.5$, the magnetic structure changes from the *C*-*E* type to the *C* type which is found at concentrations near $x=0.75$. No ideal ordering scheme for the *C*-type has been found but an arrangement can be built up of units such as that illustrated under *C*-2-6 in which one-third of the Mn^{+4} ions are mismatched with respect to the preferred antiferromagnetic coupling. It must be emphasized, however, that the *C*-2-6 entry is not now an octant of the unit cell but it is included to illustrate the type of ordering suggested.

For ion concentrations near $x=1$ the *G* type structure is stable. This structure can be considered as consisting of two interpenetrating antiferromagnetically coupled lattices. For $CaMnO_3$ samples with some admixture of Mn^{+3} ions the data (see Sec. 13) were found to be consistent with the hypothesis that the net moment of each lattice is reduced because the Mn^{+3} ions enter the lattice with opposite spin to the Mn^{+4} ions which they replace, and hence it can be concluded that in this region also the Mn^{+3}-O-Mn^{+4} exchange is ferromagnetic.

It has been found that the set of magnetic structures and transitions which have been deduced from the experimental data on this series of compounds are in most respects consistent with the simple empirical rules (a), (b), and (c) regarding the types of indirect coupling between the manganese ions in these compounds. A basis for the types of compounds encountered in this problem and consequences thereof is given by Goodenough in terms of a semicovalent theory of exchange in the following paper wherein some discussion will be found also on the relation of the superexchange mechanism of Kramers[15] and Anderson[16] and of the double exchange of Zener[17] to these perovskite-type compounds.

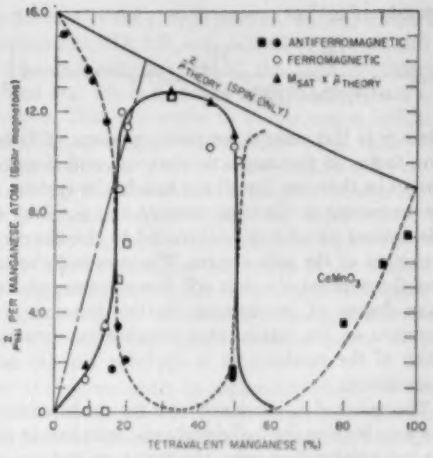


FIG. 19. Ferromagnetic and antiferromagnetic moments as a function of sample ion composition.

13. FERROMAGNETIC AND ANTIFERROMAGNETIC MOMENTS *VERSUS* COMPOSITION

The magnetic moment data which have been obtained in this investigation are summarized in Table I and are presented graphically as a function of sample composition in Fig. 19. For the discussion of this figure it will be necessary to show the relation of the plotted quantities to the experimental data.

In the neutron diffraction experiments the measured quantities are the absolute values of the integrated power in the various magnetic reflections. From the power in a given peak and the known parameters one obtains the quantities which for all the cases to be considered here[18] can be represented as $q^2jF_{hkl}^2$ where j is the multiplicity of the set of planes (hkl) and $q^2=1-(\mathbf{e}\cdot\mathbf{k})^2$ where $\mathbf{e}$ is the unit scattering vector and $\mathbf{k}$ is a unit vector in the direction of alignment of the ionic magnetic moments. For some structures, e.g., *B* and *G*, $q^2=\frac{2}{3}$ for all reflections and hence one can obtain no information about the moment orientation in the lattice. For other structures, such as Type *A*, the intensities in the pattern depend upon the moment orientation relative to the crystalline axes and therefore in such cases information relative to the moment orientation can be obtained. When q^2 has been evaluated for a given magnetic system one can obtain the quantity μ_{hkl} which is defined in terms of the measured

[15] H. A. Kramers, Physica **1**, 182 (1933).
[16] P. W. Anderson, Phys. Rev. **79**, 350 (1950); **79**, 705 (1950).
[17] C. Zener, Phys. Rev. **82**, 403 (1951).
[18] In common antiferromagnetic lattices where there are only parallel and antiparallel spin orientations q^2 can be represented as shown for unpolarized neutrons whereas in the more general case the vector $\mathbf{q}=\mathbf{e}-(\mathbf{e}\cdot\mathbf{k})\mathbf{k}$ must be appropriately introduced under the summation sign in Eq. (1).

quantity $|F_{hkl}|^2$ as

$$\mu_{hkl}=\frac{|F_{hkl}|}{n(e^2\gamma/2mc^2)f(\theta)}=\frac{1}{n}\left|\sum_n \mu_n e^{i\varphi_n}e^{2\pi i(hx_n+ky_n+lz_n)}\right|, \quad (1)$$

where γ is the neutron magnetic moment, $f(\theta)$ is the form factor of the magnetic electrons and is assumed here to be the same for all the ions in the system, μ_n is the magnitude of the ionic moment and $\varphi_n=0$ or π for orientations parallel or antiparallel to the direction of alignment of the spin system. The sum is to be taken over the n atoms of a unit cell. For a system which has some degree of randomness in the location of the moments on the lattice sites proper account must be taken of the randomness in applying Eq. (1) to the observations.

The values of μ_{hkl} obtained from the measured intensities may involve the individual ionic moments in different ways depending upon the structure and upon the particular reflection (hkl) considered. In succeeding paragraphs the measured quantities μ_{hkl}^2 are interpreted in terms of the individual ionic moments for a number of cases experimentally encountered and these are the quantities which are plotted in the figure and entered in the table. This representation has been adopted to represent on a single plot as much of the magnetic data as possible although this is not entirely satisfactory for all aspects of the problem.

For the pure A and G type structures where there is only one kind of magnetic ion, the value of μ_{hkl} as obtained from the observed F_{hkl} yields the magnetic moment per Mn^{+3} or Mn^{+4} ion respectively. It is seen that for the purest $LaMnO_3$ sample ($Mn^{+4}\sim 2\%$) the point on the graph is in good agreement with the expected value. The $CaMnO_3$ case will be considered later.

For the magnetic structure involving mixed Mn^{+3} and Mn^{+4} compositions an evaluation of the ionic moments requires in general a determination of the type of ionic ordering on the lattice sites. For the ferromagnetic cells opposite B in Fig. 18 the reflections with all indices even give $\mu_{hkl}=p_3\mu_3+p_4\mu_4$ where p_3 and p_4 are the fractional Mn^{+3} and Mn^{+4} contents, respectively, and μ_3 and μ_4 refer to the magnitudes of the moments of trivalent and tetravalent ions, respectively. This is the same result that one would obtain for a random distribution of Mn^{+3} and Mn^{+4} ions over the lattice sites. Depending upon the particular ordering scheme there will also be reflections other than those with all even indices which will yield the difference of the moments $(\mu_3-\mu_4)$. A discussion of the ordering in the ferromagnetic region was given in the last section. In any case $\bar{\mu}_F=p_3\mu_3+p_4\mu_4$ is the weighted average moment per ion in the ferromagnetic lattice and it is the square of this quantity, as determined from the neutron intensities which is plotted in Fig. 19 and listed in Table I.

In the mixed phase region ($0<x<0.25$) the measured values of $q^2jF_{hkl}^2$ for the two types of phases A and B should be proportional to $p_A\bar{\mu}_A^2$ and $p_F\bar{\mu}_F^2$, respectively, in which p_A is that fraction of the atoms in the sample which are associated with long-range antiferromagnetic ordering and p_F is the corresponding fraction for long-range ferromagnetic ordering. The average moment per ion in the ferromagnetic phase denoted by $\bar{\mu}_F$ has the definition given above. The average moment per ion in the antiferromagnetic phase $\bar{\mu}_A$ would be μ_3 if all the available Mn^{+4} ions tended to cluster in the ferromagnetic phase or would be $p_3\mu_3+p_4\mu_4$ if a fraction p_4 of the ions in the antiferromagnetic phase were tetravalent.

In the figure the effective squared moments $p_F\bar{\mu}_F^2$ and $p_A\bar{\mu}_A^2$ have been plotted as a function of ion concentration. It is evident from the figure that the growth of the effective ferromagnetic moment with increasing Mn^{+4} content in the low Mn^{+4} percentage region is less rapid than the line from 0 to $x=0.25$ which represents the maximum rate of formation of B-6-2 type ferromagnetic cells. It is not unreasonable to expect that for very small concentrations of Mn^{+4} a modified A-type structure would develop and that moderately high concentrations would be necessary for the cooperative coupling of Mn^{+3}-O-Mn^{+4} ions to produce long range ferromagnetic order. The long-range ferromagnetic order is seen to develop very rapidly near $x=0.15$, approaches full development in the vicinity of $x=0.25$ and remains large until x approaches 0.5. These results tend to lend support to the proposed ionic ordering schemes B-6-2 and B-4-4.

Information on the ferromagnetic moments, which has so far been discussed in this section in terms of the neutron measurements only, can also be obtained from direct magnetic saturation measurements (see Sec. 9). In the case of incoherent mixtures of ferro- and antiferromagnetic phases one would expect the observed saturation moment per magnetic ion in the sample to be given by $\bar{\mu}_{sat}=p_F\bar{\mu}_F$ where as previously p_F represents the fraction of the magnetic ions which are ferromagnetically aligned with an average moment $\bar{\mu}_F$. Comparison of these measurements with those obtained from neutron diffraction, which gives $p_F\bar{\mu}_F^2$, is conveniently done by plotting in Fig. 19 the quantity $\bar{\mu}_{sat}\times\bar{\mu}_F=p_F\bar{\mu}_F^2$ where the multiplying factor $\bar{\mu}_F$ is taken as the spin only value appropriate to the particular Mn^{+4} concentration. These data are represented by the triangles in Fig. 19, and it is seen that they agree quite closely with open circles from neutron diffraction measurements.

The validity of the above comparison and the apparent agreement between the neutron and saturation moment data depends on the criteria used to determine the ferromagnetic scattering from the neutron pattern. It was seen in connection with the discussion of Fig. 9 that measurements (a) above and below the Curie point and (b) with and without a magnetic field, gave essentially the same ferromagnetic scattering value in the first two reflections [but not

for the strong (222) reflection] and the moment corresponding to these two reflections agrees with that obtained from the magnetic saturation measurements.

In the case of the ferro- and antiferromagnetic mixed phases there is, however, some ambiguity in the interpretation of the ferromagnetic moments. It was shown in Fig. 10 that a magnetic field of about 4500 oersteds did not obliterate the (200) reflection and in fact for samples with low ($<14\%$)Mn^{+4} content no change in the (200) intensity was produced by the applied field. The moments determined by the field effects on the (200) reflections are represented in Fig. 19 by the squares.

If on the other hand the ferromagnetic moments are given by the total integrated intensity of the (200) temperature difference reflections one obtains the values plotted as open circles in Fig. 19. These values are seen to agree well with the corresponding ferromagnetic saturation measurements represented by the triangles in the figure.

In determining the trend of the ferromagnetic moment data the solid line was drawn through the $\bar{\mu}_F^2$ data obtained from the temperature difference curves, because it was felt that the agreement of these data with the ferromagnetic saturation measurements was strong evidence for identifying the (200) temperature difference peaks with the coherent magnetic scattering even if their intensities could not be effected by a magnetic field. The lack of field sensitivity for the coherent reflections and not for the usual saturation measurements is a point for which we have no satisfactory explanation at this time. The nature of the magnetization and hysteresis curves for the $LaMnO_3$ No. 65 sample for which the neutron magnetic field effect is negligible will also need further study before it can be satisfactorily accounted for.

For the *C-E* type structure observed at a Mn^{+4} content of about 58% the values of μ_{hkl}^2 obtained from the data can, as has been described, lead to an estimate of the moments occupying the several types of lattice sites. For the ordering shown in Fig. 18 and for the ideal composition, there will be two values of μ_{hkl}^2, of which one is proportional to μ_3^2 the other to μ_4^2. The departure of the sample from the ideal stoichiometric composition was shown to introduce modifications in the intensity pattern which could be accounted for if the excess Mn^{+4} ions were introduced into the Mn^{+3} sites at random but with reversed spin. In this case the observed μ_{hkl}^2 for the Mn^{+3} sites can be said to be proportional to $\bar{\mu}_3^2$ where $\bar{\mu}_3 = p_3\mu_3 - p_4\mu_4$ and is the weighted average, taking into account the spin reversal, of the moments of the ions on the ideal Mn^{+3} sites. For this case one cannot define an average magnetic moment in the sense described above and hence no entry is made in the figure.

In the case of the *C* type pattern, the μ_{hkl} values for reflections 110, 112, 130, etc., will all be proportional to $p_3\mu_3 + p_4\mu_4 = \bar{\mu}_A$ irrespective of the type of ordering provided that where there are misfits there are no spin inversions. That this seems to be the case is indicated by the fair agreement of the observed $\bar{\mu}$ value with that expected from the above model (2.99 observed *vs* 3.20 calculated). The observed value is however somewhat lower than expected and may represent an incomplete development of the structure. Accordingly for this case also no entry has been made in the figure.

The suggested growth of antiferromagnetic phases in the region $x>0.5$ is indicated by a rising and leveling off of the antiferromagnetic moment curve in this region.

The *G* type structure is apparently stable for compositions with $x>0.8$. For this structure only reflections which have all indices odd are permitted and for them $\mu_{hkl} = p_4\mu_4 \pm p_3\mu_3 = \bar{\mu}_A$. The + sign is to be taken if the Mn^{+3} ion in compositions different from $x=1$ replace Mn^{+4} with the same spin and the negative sign if they go into the lattice with spin opposite to the replaced Mn^{+4} ions. The predicted curve for this latter case is shown in the figure and it seems to agree well with the observed points. For the first case the average moments would have increased with Mn^{+3} content.

14. ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance received from their colleagues in the laboratory. They are indebted to Mr. R. H. Ward for assistance in the design of the spectrometer, to Mr. D. E. LaValle for assistance with sample preparations, to Mr. A. D. Horton for the chemical analyses and to Dr. H. L. Yakel for his x-ray studies and permission to use his data prior to publication. Helpful conversations and correspondence with Dr. John B. Goodenough of Lincoln Laboratories regarding the theoretical aspects of the problem are also gratefully acknowledged.

Note added in proof.—Additional data on the temperature dependence of lattice constants similar to that shown in Fig. 17 has recently been obtained by Yakel. These data concern the nearly ionically pure sample of $LaMnO_3$ No. 12 which shows the large lattice distortion at room temperature and below. The new data taken at elevated temperatures show a slow gradual decrease in splitting up to about 400°C after which there is a much more rapid decrease until at about 485°C the apparent symmetry becomes cubic. This constitutes a further illustration of the ordering transformation in the directed orbitals proposed by Goodenough for these compounds. We wish to thank Dr. H. L. Yakel for permission to quote these results.

Theory of the Role of Covalence in the Perovskite-Type Manganites $[La,M(II)]MnO_3$†

JOHN B. GOODENOUGH
Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts
(Received May 16, 1955)

The theory of semicovalent exchange is reviewed and applied to the perovskite-type manganites $[La,M(II)]MnO_3$. With the hypothesis of covalent and semicovalent bonding between the oxygen and manganese ions plus the mechanism of double exchange, detailed qualitative predictions are made about the magnetic lattice, the crystallographic lattice, the electrical resistivity, and the Curie temperature as functions of the fraction of Mn^{4+} present. These predictions are found to be in accord with recent findings from neutron-diffraction and x-ray data as well as with the earlier experiments on this system by Jonker and van Santen.

I. INTRODUCTION

IN a recent paper[1] on the influence of covalent effects in spinel-type structures, it was shown how covalent bonding influences lattice distortions and indirect magnetic-exchange interactions between two magnetic cations separated by an anion. The magnetic interaction was termed semicovalent exchange since it depends upon the concept of semicovalent bonding, a concept which was also introduced in that same paper. Although there are certain similarities between this exchange mechanism and the superexchange mechanism which was first proposed by Kramers[2] in 1934, the two have fundamental differences: in some oxides the two lead to identical predictions, but in others they lead to opposite predictions. The perovskite-type manganites represent a system in which the magnetic moments of the manganese ions are coupled by indirect exchange mechanisms; it is a system for which semicovalent exchange and superexchange give opposite predictions. It is, therefore, a particularly appropriate system for evaluating the merits of the semicovalent-exchange hypothesis, especially as Wollan and Koehler,[3] with the support of x-ray work by Yakel, have just completed a beautiful experimental study of its magnetic and crystallographic lattices as functions of Mn^{4+} content.

In Sec. II, there is a discussion of semicovalence, and in Sec. III there is a restatement of the indirect magnetic coupling between the various manganese ions in a perovskite-type oxide if semicovalence occurs. The effects of covalence on bond length are also pointed out. This discussion is summarized in Table I. The predictions resulting from the superexchange model are contrasted with those from the covalent-bond and semicovalent-exchange models which, in Sec. IV, are applied to the system $[La,M(II)]MnO_3$, where $M(II)$ is Ca^{2+}, Sr^{2+}, or Ba^{2+}. These predictions are summarized in Sec. V and Fig. 6.

† The research in this document was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.

[1] J. B. Goodenough and A. L. Loeb, Phys. Rev. **98**, 391 (1955).

[2] H. A. Kramers, Physica **1**, 182 (1934); P. W. Anderson, Phys. Rev. **79**, 350 (1950); J. H. Van Vleck, J. phys. radium **12**, 262 (1951).

[3] E. O. Wollan and W. C. Koehler, preceding paper [Phys. Rev. **100**, 545 (1955)].

II. SEMICOVALENCE

There are three principal contributions to the internal energy of a solid, the electrostatic, electronic, and elastic contributions. In an ideal ionic lattice the electronic contribution is negligible, the principal binding energy coming from the electrostatic term. In a homopolar or metallic lattice the electrostatic term is negligible, the principal binding energy coming from the electronic term.

Previous workers have assumed that the transition-element cations in oxide lattices are ionically bound. Although the superexchange mechanism depends upon the nonionic component of the cation-oxygen bonds, in this model it is presumed that the atomic electron orbitals provide a fairly accurate description for the electrons in the lattice, or that a tight-binding approximation is valid. Whereas various authors[4] have considered the perturbations of the cation d orbitals caused by internal electric fields, there has been no serious consideration of the perturbations of the empty cation orbitals. The empty cation orbitals are assumed to have so much higher an energy than either the cation d orbitals or the anion p orbitals that their influence on the state of the electrons in the system is qualitatively similar to that of unperturbed empty orbitals.

Such an assumption is extremely dangerous. Because of the strong perturbations of neighboring atoms on one another, the empty energy levels correspond to lattice orbitals, not atomic orbitals, and these lattice orbitals may have energies which are nearly degenerate with the atomic d orbitals. A qualitative table has been given[1] in which the most stable hybrid orbitals were listed for various transition-element cations. Because of extreme steric hindrance in a close-packed lattice, the most stable of the cation hybrid orbitals may not overlap appreciably the neighboring anion orbitals. In such a case the empty lattice orbital may have little effect on the state of the electrons in the solid, and the assumption of ionic bonding is suitable. However, if the most stable of the empty cation orbitals strongly overlap the full orbitals of neighboring anions, the anion p electrons may spend some of their time in the cation orbitals. The

[4] H. A. Bethe, Ann. Physik **3**, 133 (1929); J. H. van Santen and J. S. van Wieringen, Rec. trav. chim. **71**, 420 (1952).

amount of electron sharing between cation and anion depends, of course, on the relative stabilities of their orbitals. The degree of electron sharing is a measure of the homopolar bonding.

The full anion p orbital which overlaps an empty cation orbital contains two electrons of opposite spin. In ordinary homopolar bonding each of these electrons has an equal probability of being shared by the cation. If the cation has an oriented net magnetic moment, however, these electrons will not have an equal probability of being shared by the cation. Because of the presence of exchange forces, that anion electron whose spin is parallel to the net cation spin will spend more time on the cation than that with antiparallel spin. This is in accordance with Hund's rule if it is imagined that the d shell is extended by the s and p orbitals when an atom is in a lattice where hybrid orbitals are formed. Because a single electron predominates in this bond, it is called semicovalent.

Semicovalence assumes, therefore, that the exchange energy between the shared electron and the cation d-shell electrons is greater than the energy difference between the empty lattice orbital and the d-orbital energy levels occupied by one electron. It can only occur in those bonds where there is strong overlap of stable, empty cation orbitals and an anion p orbital; semicovalent bonds are directed in the same manner as normal covalent bonds. Semicovalence occurs only below the Curie temperature where the net cation magnetic moment is oriented. Above the Curie temperature there is normal coordinate covalence. If the formation of covalent bonding causes a distortion of the lattice from cubic symmetry, there is a transition temperature above which the covalent bonding is not ordered and cubic symmetry results. This transition temperature is above the Curie temperature if the magnetic coupling is due to semicovalent exchange. Normally a distortion of a lattice from cubic symmetry increases the energy gap between the valence and conduction bands so that covalent-bond ordering is a cooperative phenomenon[1] and the ordering temperature can be defined within fairly narrow temperature limits. However, it is possible that covalent-bond ordering may, in some instances, reduce the energy gap between the conduction band and the valence band. In such a case there will not be a well-defined ordering temperature; rather, covalent-bond ordering will occur over a wide temperature range.

III. Mn–Mn MAGNETIC EXCHANGE

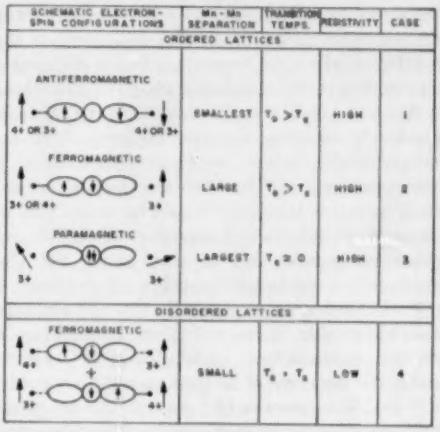
In order to give an interpretation of the system $[La,M(II)]MnO_3$, it is first necessary to outline the model of Mn–O bonding and Mn–Mn magnetic exchange on which this interpretation is based.

First it is assumed that Mn^{3+}, with an outer-electron configuration d^4, hybridizes stable ($ds p^2$) lattice orbitals. These lattice orbitals are square and coplanar. The reason for postulating their stability is the availability of but a single empty d orbital. Experimental evidence[1] for this postulate is the tetragonality of hausmannite, Mn_3O_4, with $c/a>1$. To be consistent with this assumption, it is further postulated that Mn^{4+}, with outer-electron configuration d^3, hybridizes stable (d^2sp^3) orbitals, two empty d orbitals being available. These lattice orbitals are octahedral and can therefore point simultaneously toward all six oxygen near neighbors if the Mn^{4+} ion is in an octahedral site, as it is in the perovskite-type manganites $[La,M(II)]MnO_3$. The preference of Cr^{3+} for octahedral coordination when in a spinel lattice is experimental evidence that cations with outer-electron configurations d^3 do hybridize (d^2sp^3) orbitals which are stable enough to effect the cation-anion bonds.[1,5]

In the perovskite-type manganites there are three possible Mn–O bonds: a covalent or semicovalent bond if an empty manganese orbital points toward the O^{2-} ion, an ionic bond if the empty Mn^{3+}–($ds p^2$) orbitals point away from the O^{2-} ion, or a metallic-like bond if the O^{2-} ion is between a Mn^{3+} and a Mn^{4+} ion in a lattice of disordered Mn^{3+}, Mn^{4+} ions. The first of these bonds is the most stable; it is therefore the shortest bond. The second is the least stable; it is the longest

TABLE I. Semicovalent model for the magnetic coupling of manganese ions in the perovskite-type manganites $[La,M(II)]MnO_3$. N is the number of semicovalent bonds which can be formed by a manganese ion with its six neighboring oxygen ions. T_θ and T_c are the transition temperatures for bond ordering and magnetic ordering, respectively. In the column for schematic electron-spin configurations the cations, marked 4+ and 3+, have an empty orbital pointing toward the O^{2-} p orbitals if they are joined by a dash.

Ion	Outer elec. config.	Empty low-energy orbitals	N
	Low-energy, empty lattice orbitals		
Mn^{4+}	d^3	Octahedral (d^2sp^3)	6
Mn^{3+}	d^4	Square ($ds p^2$)	4



[5] E. W. Gorter, Philips Research Repts. 9, 403 (1955).

bond. These bond types are admixed above some transition temperature. The transition from covalent or ionic bonding to admixed bonding is not necessarily associated with a magnetic transition. Judging from the covalent transition[1] in Mn_3O_4 at 1170°C, the covalent-admixed transition should occur well above room temperature in the perovskite-type manganites. Any ionic-admixed transition temperature should also be greater than the Curie temperature. On the other hand a transition from semicovalent to covalent, or from metallic-like to admixed, bonding is accompanied by a magnetic transition. If two or more bond types occur in the same crystal, the long-range magnetic transitions all occur at about the same temperature because of the cooperative character of magnetic ordering.

There are four possible Mn−O−Mn bonding arrangements, and with each of these four arrangements there is a different Mn−Mn separation and a different indirect coupling of the manganese magnetic moments. If both Mn−O bonds can be covalent, the Mn−Mn separation is smallest and the cations couple antiferromagnetically below the Curie temperature, the transition temperature from two semicovalent bonds to partially covalent bonds. It should be noted that it is the same O^{2-} p orbital which extends towards both near-neighbor manganese ions so that above the Curie temperature no one bond is covalent all the time; the two O^{2-} electrons are shared equally with the manganese ions on either side. Below the Curie temperature the bonds are semicovalent, each of the O^{2-} electrons spending the majority of its time on a different side of the O^{2-} ion. Since the two O^{2-} electrons are antiparallel and, by semicovalence, each couples ferromagnetically with the net manganese-ion spin with which it is predominantly associated, the two manganese ions are coupled antiferromagnetically. The Curie temperature is a measure of the exchange-energy gain associated with the ordering of the anion electrons. This argument is summarized in Case 1 of Table I.

If only one Mn−O bond can be covalent, below the covalent-admixed transition one bond is ionic, the other covalent. The Mn−Mn separation is greater than in the case where both bonds can be covalent, and the O^{2-} ion is displaced toward the manganese ion with which it forms the covalent bond. Below the Curie temperature the covalent bond becomes semicovalent so that a net magnetic moment is associated with the O^{2-} ion. This moment is due to the anion electron which participates less in the semicovalent bond; it is therefore antiparallel to the net moment of the manganese ion which is coupled ferromagnetically via semicovalent bonding to the other anion electron. Since there is a direct exchange interaction between the magnetic moment of the anion and that of the neighboring manganese ion which is ionically bound to it, the second manganese-ion moment is also antiparallel to the anion moment, and the two manganese ions are coupled ferromagnetically, as shown in Case 2 of Table I.

If neither Mn−O bond can be covalent, both bonds are ionic. The Mn−Mn separation is largest, and these is no indirect magnetic exchange between the manganere ions. This is illustrated as Case 3 of Table I

Finally, if the anion separates ferromagnetically coupled Mn^{3+} and Mn^{4+} ions in a lattice of disordered Mn^{3+}, Mn^{4+} ions, the state $Mn^{3+}-O-Mn^{4+}$ is degenerate with the state $Mn^{4+}-O-Mn^{3+}$ so that below the Curie temperature ferromagnetic double exchange takes place between the manganese ions, one electron being free to jump through the anion between the two manganese ions to form a metallic-like bond. The double-exchange mechanism was first proposed by Zener.[6] If the Mn^{3+} and Mn^{4+} ions form part of an ordered lattice, however, double exchange is inhibited as the degeneracy is removed. It should also be noted that low electrical resistivity accompanies the double-exchange phenomenon; high electrical resistivity is associated with the other exchange mechanisms even if there is ferromagnetic coupling of the manganese magnetic moments.

These predictions are in sharp contrast to those based on a superexchange model. Although superexchange is also based on the nonionic character of the lattice, it is assumed in this model that the cation s and p orbitals are of little relevance; the electron which is excited from an anion to a neighboring transition-element cation is said to go into a cation d orbital. According to this model it is that anion electron which is antiparallel to the net cation spin which is excited whenever the d shell is half or more filled, that electron which is parallel whenever the d shell is less than half full. Also in the superexchange model only one anion electron can be excited at a time so that two magnetic cations on opposite sides of an anion are coupled parallel if their d shell is less than half full, antiparallel if the d shell is half or more filled. In this model it is assumed that the exchange energy is smaller than the energy difference between the lattice-orbital and the d-orbital energy levels whereas the opposite is assumed in the semicovalent-exchange model. From such a model the system $[La,M(II)]MnO_3$ should be ferromagnetic for all compositions with a maximum Curie temperature and a minimum electrical resistivity for $x=0.5$, where x is defined as the fraction of manganese ions which are Mn^{4+}. This follows from the fact that both Mn^{3+} and Mn^{4+} ions have d shells which are less than half full and that double exchange should be a more stable mechanism than superexchange. This prediction is not verified by experiment: the ferromagnetic compositional region is only found for $0.2<x<0.4$, and the maximum Curie temperature and minimum electrical resistivity is in the composition $x\approx0.3$.[7] In the superexchange model there is no way for atoms to couple ferromagnetically in one direction, antiferromagnetically in another as it con-

[6] C. Zener, Phys. Rev. **82**, 403 (1951).

[7] G. H. Jonker and J. H. van Santen, Physica **16**, 337 (1950); J. H. van Santen and G. H. Jonker, Physica **16**, 599 (1950).

tains no anisotropy; similarly the superexchange model cannot account for lattice distortions. Wollan and Koehler[3] have shown that both of these features are required of an adequate model for the system $[La,M(II)]MnO_3$. It can be concluded that superexchange is completely inadequate for this chemical system; it remains to be shown that the covalent and semicovalent-exchange model is adequate.

IV. INTERPRETATION OF THE PEROVSKITE-TYPE SYSTEM $[La,M(II)]MnO_3$

Attention is first focused on the manganese ions of the manganites as these are believed to be primarily responsible for the lattice distortions, magnetism, and electrical-resistivity variations which occur. Discussion of the large-cation effects are deferred until the end of this section. Furthermore, it is assumed that the lowest internal energy of a system is that which permits the largest number of covalent, or semicovalent, manganese-oxygen bonds. Frequent reference will be made to the experimental determinations of Wollan and Koehler.[3] The fraction of manganese ions that are Mn^{4+} is designated by x, and a perovskite-type lattice with manganese ions in octahedral O^{2-} interstices is assumed throughout.

A. Compositions $x=1$

In the ideal case of $x=1$, all the manganese ions are Mn^{4+} and every Mn−O bond can be semicovalent. The $Mn^{4+}-Mn^{4+}$ coupling corresponds, therefore, to Case 1 of Table I, and each Mn^{4+} ion is antiferromagnetically coupled to its six near-neighbor cations as shown in Fig. 1. This is just the magnetic lattice which was observed by Wollan and Koehler for $CaMnO_3$ with $x=0.97$. It corresponds to their type-G magnetic lattice. Since all the Mn−O bonds are identical, cubic symmetry is predicted for the lattice. This was also observed.

B. Compositions $0.9<x<1$

If one of the Mn^{4+} ions of a composition with $x=1$ is replaced by a Mn^{3+} ion, the Mn^{3+} ion is surrounded by Mn^{4+} ions of opposite spin to the one it replaced. Since a

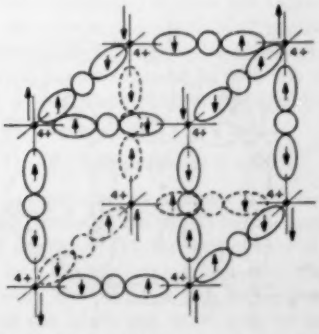


FIG. 1. Magnetic lattice for $x=1$; type G.

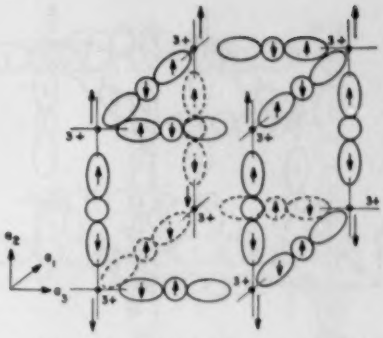


FIG. 2. Magnetic lattice for $x=0$; type A.

single Mn^{3+} ion cannot be said to be ordered, it couples ferromagnetically with its six neighbors via double exchange, Case 4 of Table I. (However, because of the antiferromagnetic character of the parent lattice, the extra electron associated with the Mn^{3+} ion does not migrate freely through the lattice as would a single electron in the conduction band of a semiconductor). The generalization then follows that if a small fraction of Mn^{3+} ions are present, these ions should orient themselves antiparallel to the Mn^{4+} ion they replace; this reduces the net moment of the magnetic sublattices to which they belong. The Mn^{3+} ions are randomly distributed, so the two sublattices of the type-G structure are equally reduced; Wollan and Koehler have found this effect by measuring a decrease in the average moment per ion in the antiferromagnetic phase with increasing Mn^{3+} content (see their Fig. 19). This effect should predominate in the compositional region $0.9<x<1$ since the Mn^{3+} ions are essentially isolated and randomly distributed at these low Mn^{3+} concentrations. Further, since the Mn−O bonds of the double-exchange model are isotropic, the cubic structure of the lattice is preserved. Finally, since the double-exchange mechanism is stronger in this system than semicovalent exchange, the Curie temperature may be expected to increase with increasing Mn^{3+} content.

C. Compositions $x=0$

In an ideal composition with $x=0$, all the manganese ions are Mn^{3+} and only $\frac{3}{4}$ of the Mn−O bonds can be covalent or semicovalent. The covalent bonds order below some transition temperature in order to reduce the elastic energy associated with the different Mn−Mn separations associated with different bond types. This ordering produces a cooperative elastic strain within the crystal. One possible arrangement of these bonds is an ordering of all the coplanar bonds in (100) planes. In such an arrangement the Mn−O−Mn bonds within a (100) plane correspond to Case 1 of Table I whereas those between (100) planes correspond to Case 3. Such an ordering would produce a large crystalline distortion

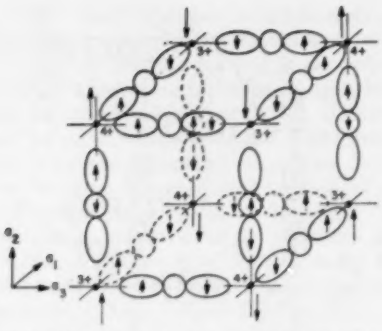


FIG. 3(a). High-energy ordered structure for $x=0.5$; type C.

to tetragonal symmetry with axial ratio $c/a>1$: a similar type of ordering in the spinel lattice[1] of Mn_3O_4 produces an axial ratio $c/a=1.16$. It is immediately seen that in the perovskite structure there is another ordered configuration for the covalent bonds which produces a smaller lattice distortion and is therefore more stable, less energy being expended to distort the cubic symmetry. This configuration is pictured in Fig. 2. In this arrangement all of the O^{2-} ions are bonded covalently in at least one direction: in the (100) planes perpendicular to the a_2 axis the Mn−O−Mn bonds all correspond to Case 2 of Table I whereas all the bonds parallel to the a_2 axis correspond to Case 1. Such an ordering of the covalent bonds produces a lattice distortion with $a_2<a_1=a_3$: however, this distortion is smaller than the former possibility discussed because the Mn−Mn separation in Case 2 is smaller than that in Case 3. If use is made of the axial ratio of Mn_3O_4 and of the lattice parameters $a=8.14$ A, $a_2=7.68$ A of Mn_3O_4 and $LaMnO_3$, respectively, the strain $\Delta\equiv(a_3-a_2)/a_3=0.042$ is calculated.

Further, it was pointed out in the discussion of Case 2 of Table I that the O^{2-} ion is closer to the manganese ion with which it forms a covalent bond than it is to its other neighbor with which it forms an ionic bond. This means that in the planes perpendicular to the a_2 axis the oxygen ions about any Mn^{3+} ion form a rhombus rather than a square. Now a characteristic feature of perovskite-type lattices is a "puckering" of the oxygen octahedra so that the unit cell is double that of the simple perovskite-type structure. This puckering is found, for example, in perovskite itself, $CaTiO_3$.[8] Since the O^{2-} ions in the manganese-containing planes perpendicular to the a_2 axis are bonded covalently to only one of their near neighbors, there is no distortion of covalent-bond angles if "puckering" occurs in this plane. Whereas such a puckering would not destroy the square symmetry of the plane were the manganese ion at the center of a square formation of O^{2-} ions, after puckering the square symmetry is replaced by rhombic symmetry when the manganese ions are at the center of a rhombic formation of O^{2-} ions and the near-neighbor rhombuses alternate the direction of their long axis, as in the case corresponding to Fig. 2. It follows, therefore, that at $x=0$ the lattice may be orthorhombic, rather than tetragonal, with $a_2<a_3=a_1$. The x-ray data reported by Wollan and Koehler in a sample with $x=0.02$ indicates orthorhombic symmetry and $a_2<a_3=a_1$ with the ratio $a_2/a_3=0.965$. An extrapolation of their values to $x=0$ gives the ratio $a_2/a_3=0.959$ with $\Delta=0.041$.

Finally, if the covalent bonds order as in Fig. 2, below the Curie temperature the magnetic coupling within the (100) planes is ferromagnetic, but that between planes is antiferromagnetic, Cases 2 and 1, respectively, of Table I. But this is just the type-A magnetic lattice which is reported by Wollan and Koehler.

It should be noted that the transition temperature for covalent-bond ordering may be much higher than the Curie temperature. The crystalline distortion from cubic symmetry occurs at the transition temperature for covalent-bond ordering; at the Curie temperature the change in lattice parameter is only a second-order effect.

D. Compositions $0<x<0.1$

If a Mn^{4+} ion replaces one of the Mn^{3+} ions of the type-A lattice illustrated in Fig. 2, all six, instead of just four, of the oxygen ions around that site can form covalent bonds. Therefore the addition of Mn^{4+} ions reduces the ratio a_2/a_3 and the transition temperature for covalent-bond ordering is lowered. Further, below the Curie temperature the spin of the Mn^{4+} ion remains parallel to that of the Mn^{3+} ion it replaced since four of the six bonds favor similar magnetic coupling to that with only Mn^{3+} ions present: Mn^{4+}, which is randomly located, couples ferromagnetically with its four neighbors in the (100) plane via double exchange, Case 4. Also it is capable of coupling antiferromagnetically with its two neighbors along the a_2 axis via semicovalent exchange, Case 1. Thus below the Curie temperature the Mn^{4+} ion is constrained to stay in its (100) plane, the antiferromagnetic coupling between planes inhibiting electron transfer from one plane to the other.

It is concluded that the type-A antiferromagnetic lattice should predominate in the compositional region $0<x<0.1$ since the Mn^{4+} ions are essentially isolated

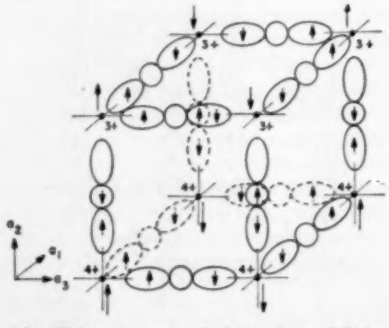


FIG. 3(b). High-energy magnetic lattice for $x=0.5$; type C.

[8] H. D. Megaw, Acta Cryst. 5, 739 (1952).

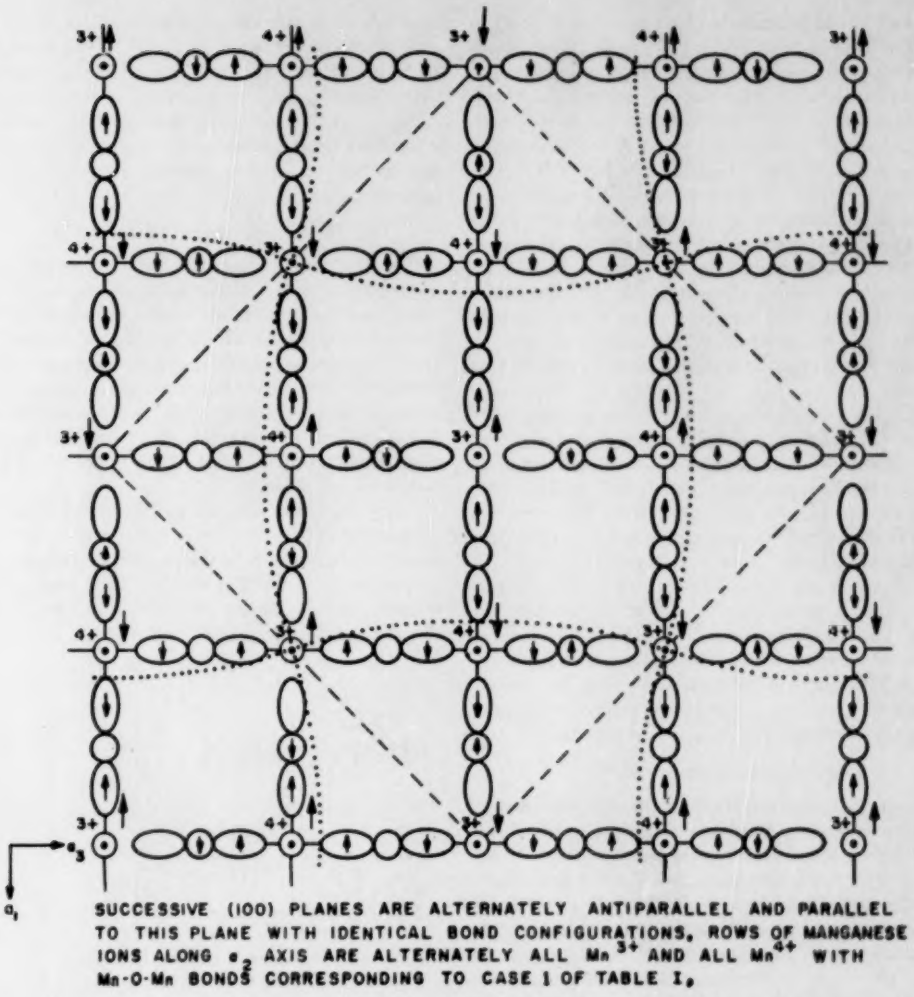

3+
4+
a_3
a_1
SUCCESSIVE (100) PLANES ARE ALTERNATELY ANTIPARALLEL AND PARALLEL TO THIS PLANE WITH IDENTICAL BOND CONFIGURATIONS. ROWS OF MANGANESE IONS ALONG a_2 AXIS ARE ALTERNATELY ALL Mn^{3+} AND ALL Mn^{4+} WITH Mn-O-Mn BONDS CORRESPONDING TO CASE 1 OF TABLE I.

FIG. 4. Low-energy magnetic lattice for $x=0.5$; type *CE*.

and randomly distributed at these low values of x. The Curie temperature increases with Mn^{4+} content because of the added possibilities for double exchange. The axial ratio a_2/a_3 increases toward one with the number of bonds in the a_1, a_3 plane which are changed from Case 2 to Case 4 of Table I.

E. Compositions $x=0.5$

If $x=0.5$, there are just as many Mn^{3+} ions as Mn^{4+} ions present. Although the M(II) ions or lattice vacancies which are responsible for the presence of the Mn^{4+} ions may be randomly located, the ready migration of electrons from Mn^{3+} to Mn^{4+} ions makes it possible for ordering of the Mn^{3+} and Mn^{4+} ions to occur. This ordering occurs at transition temperatures which are similar to the covalent-bond ordering temperature since the driving force behind this ordering is due more to covalent-bond formation than Madelung-energy considerations.

At $x=0.5$ there are three possible ordered-bond arrangements which give cooperative lattice strains: these are pictured in Figs. 3(a), 3(b), and 4. They correspond to an ordering of Mn^{3+} and Mn^{4+} ions in alternate (111), (100), or (110) planes, respectively. The electrostatic energies for these three configurations are only slightly different; the elastic energy, however, is markedly reduced for the configuration of Fig. 4. This follows from the fact that in the configurations in Fig. 3 the ratio $(a_3-a_2)/a_3$ corresponds to that for the composition $x=0$, whereas in Fig. 4 this ratio is only half as great

since half of the bonds in a (100) plane correspond to Case 1 of Table I. The ordering within the (100) planes gives the optimum electrostatic configuration as well as a cooperative elastic strain which conserves the square symmetry. If the Mn^{3+} ions establish the square symmetry, the Mn^{4+} ions are displaced in the cooperative way indicated by the dotted lines in Fig. 4. Further, there are sufficient covalent bonds in these planes to inhibit a puckering of the oxygen octahedra. From binding-energy considerations it is predicted, therefore, that the covalent bonds order as shown in Fig. 4 and, using the lattice parameters for $x=0$, that the lattice distortion from cubic symmetry is to tetragonal with $c/a=0.979$. This corresponds to a measured value, as reported by Wollan and Koehler, of $c/a=0.974$ at $x=0.58$.

Below the Curie temperature semicovalent bonds are formed to produce completely cooperative magnetic coupling between the manganese ions: the resulting magnetic lattice is just the type-*CE* configuration which was determined by Wollan and Koehler from neutron-diffraction experiments on a sample with $x=0.58$. It is illustrated in Fig. 4.

Finally, it is suggested that if the high-temperature cubic phase is puckered, the energy gap between the conduction band and the valence band might be reduced by a bond-ordering which inhibited a puckering of the lattice. This could account for the lack of a sharp ordering temperature on going from tetragonal to cubic symmetry as reported by Wollan and Koehler.

F. Compositions $x=0.75$

At $x=0.75$, half of the Mn^{3+} ions in any (100) plane of Fig. 4 are replaced by Mn^{4+} ions, and a cooperative elastic strain within the plane can only occur if Mn−O−Mn bonds corresponding to Case 2 of Table I are aligned along the same axis. This is in contrast with the type-*CE* lattice shown in Fig. 4 where the Case-2 Mn−O−Mn bonds are so ordered that square symmetry within the (100) planes is preserved without requiring any Mn−O−Mn bond to be strained. This alignment of the ionic Mn−O bonds means that along two axes all bonds correspond to Case 1 whereas along the third axis half of the bonds correspond to Case 2. Such an ordering of the covalent bonds gives an axial ratio $c/a=1.020>1$ if the lattice parameters for $x=0$ are used to give a measure of the relative Mn−Mn separations for Cases 1 and 2. The measured[3] axial ratio for a sample with $x=0.8$ is $c/a=1.022$. It should be noted that elastic considerations inhibit the location of neighboring Mn^{3+} ions along the c axis as such a configuration would introduce a Mn−Mn configuration corresponding to Case 3. If the ions order as shown in Fig. 5, the elastic energies are minimized; each row of manganese ions parallel to the c-axis have the same length.

Below the Curie temperature the magnetic exchange interactions are not completely cooperative. Because of the ordering induced by elastic-energy considerations, there are no double-exchange interactions. If the magnetic lattice orders as shown in Fig. 5, one third of the Mn^{4+} ions have only two of their six near-neighbor interactions opposed to the configuration while four favor it, and all other interactions are cooperative. This is the most nearly cooperative coupling possible, and it corresponds to Wollan and Koehler's type-*C* configuration.

G. Compositions $0.5<x<0.75$

If a Mn^{4+} ion replaces a Mn^{3+} ion in the type-*CE* configuration of Fig. 4, the lattice distortion is diminished since the number of covalent bonds in the (100) plane is increased. Also the Mn−Mn separations in the (100) plane are no longer completely cooperative and the covalent-bond ordering temperature is reduced. Experimentally it is observed that the covalent-ordering transition-temperature region decreases from just above room temperature to below room temperature in this compositional range.

Below the Curie temperature the added Mn^{4+} ion is antiparallel to the Mn^{3+} ion it replaces since four out of six of its neighbors favor this magnetic arrangement: it couples ferromagnetically with the Mn^{3+} ions in the two neighboring (100) planes via double exchange, Case 4 of

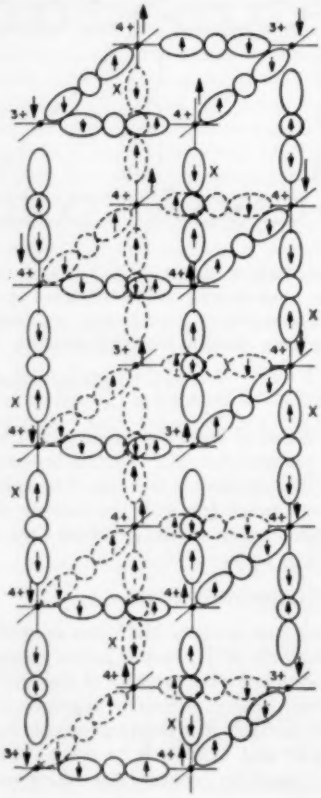


FIG. 5. Magnetic lattice for $x=0.75$; type *C* bonds marked with an *X* are mismatched.

Table I, while the four Mn^{4+} neighbors compete equally via Case 1 for a spin parallel or antiparallel to that of the original Mn^{3+} ion. This corresponds to the experimental conclusions of Wollan and Koehler from neutron-diffraction intensity measurements in patterns from compositions with $0.5<x<0.71$.

At some point between $x=0.5$ and $x=0.75$ there will be a two-phase region in which the modified type-*CE* elastic and magnetic configuration described above is mixed with the type-*C* pattern corresponding to $x=0.75$. This is probably an extremely small region near the composition $x=0.75$ since ordering of the bonds within the (100) planes should be a cooperative phenomenon.

H. Compositions $0.75<x<0.9$

If a Mn^{4+} ion replaces a Mn^{3+} ion in the type-*C* lattice of Fig. 5, the lattice distortion and the covalent-bond transition temperature are reduced; magnetically the Mn^{4+} ion couples parallel to the Mn^{3+} ion it replaces, and the type-*C* magnetic structure is preserved. As more Mn^{4+} ions are added, however, a composition is reached where the elastic forces no longer order the lattice; the structure then becomes cubic with a magnetic configuration corresponding to the modified type-*G* lattice described for compositions $0.9<x<1$. Between the type-*C* and the modified type-*G* lattices there is a two-phase region: this region probably lies within the compositional range $0.85<x<0.9$.

I. Compositions $0.25<x<0.375$

At $x=0.25$ the electrostatic energy is minimized if the Mn^{4+} ions order so as to form a b.c.c. array of cell edge two Mn−O−Mn distances. A subsequent ordering of the covalent bonding which makes every Mn^{4+}−O−Mn^{3+} bond correspond to Case 4 of Table I and every Mn^{3+}−O−Mn^{3+} bond to Case 2 would produce a cooperative elastic strain to rhombohedral symmetry, the Case 4 Mn−Mn separations being smaller than those corresponding to Case 2. Further, there is ferromagnetic coupling throughout and a mechanism for relatively high electric conductivity through the Case 4 bonds.

In such a structure the strength of the double-exchange mechanism is reduced by the electrostatic ordering, the reduction increasing with the polarity difference between the ions designated Mn^{4+} and Mn^{3+}. It is probable, therefore, that at last at higher temperatures the Mn^{4+} ions are randomly distributed throughout the lattice so that the lattice remains cubic even below the Curie temperature where the prevalence of double exchange makes the lattice ferromagnetic.

For the compositional range $0.25<x<0.375$ there is less reason to expect an ordering of the Mn^{4+} ions. The disordered lattice is cubic with a relatively high Curie temperature. Since its bonds are metallic-like, it has a low resistivity. The maximum Curie temperature and minimum electrical resistivity must occur at that composition for which the double-exchange mechanism is optimized; this optimum occurs for that disordered composition which has the largest number of Mn^{3+} ions with one, and only one, Mn^{4+} near neighbor. In an ordered lattice this occurs for $x=0.25$; in a disordered lattice this occurs at an $x>0.25$. Since the double-exchange mechanism requires a disordered arrangement, in the Appendix there is a calculation of the value of x at which there are the largest number of Mn^{3+} ions with one, and only one, Mn^{4+} near neighbor in a random matrix of Mn^{4+} and Mn^{3+} ions. This value is $x=0.31$. Any deviation of the lattice from complete randomness of the Mn^{4+} ions reduces this value toward $x=0.25$. The measurements of Jonker and van Santen[7] indicate a maximum Curie temperature and a minimum resistivity at $x\approx 0.3$.

J. Compositions $0.1<x<0.25$ and $0.375<x<0.5$

The compositional regions between the modified type-*A* and the ferromagnetic phases and between the ferromagnetic and the type-*CE* phases must be two-phase. Therefore within $0.1<x<0.25$ there must be a two-phase region in which the modified type-*A* and the ferromagnetic lattice are both present. The axial ratio a_2/a_3 for the solubility limit of the modified type-*A* lattice will be closer to one than its value at $x=0$; nevertheless it appears it should be possible to detect the two-phase character of this lattice by x-ray as well as neutron-diffraction methods as one phase is cubic, the other is not. Wollan and Koehler have observed the two magnetic phases; they were unable to detect two crystallographic phases within this compositional region.

Similarly within the compositional range $0.375<x<0.5$ there must be a ferromagnetic, cubic phase coexistent with an antiferromagnetic, tetragonal phase. Again only the two magnetic phases have been identified.

K. Effect of the Large Cations

So far only the role of the manganese cations has been considered. The large cations must also influence the lattice structure and the covalent-bond formation. If the bivalent cations which replace the La^{3+} ions are too large, there will be a composition at which the perovskite structure is no longer maintained. This is found[7] for (La,Sr)MnO_3 and (La,Ba)MnO_3 at concentrations of Sr^{2+} or Ba^{2+} much above 0.5. The series (La,Ca)MnO_3 forms solid solutions of perovskite-type structure throughout the entire compositional range.

If the large cations form bonds with considerable covalent character in this structure, this influences the transition temperature from an ordered to disordered covalent-bond arrangement. The most stable hybrid orbitals of the larger cations are tetrahedral (sp^3) orbitals. If the large cations form partially covalent bonds with four of their neighboring twelve oxygen ions, these bonds are in planes which are perpendicular to the edges of the cubes formed by the manganese ions. If a p orbit of an O^{2-} ion is oriented through covalence toward a large cation, it has a perpendicular p orbit which may

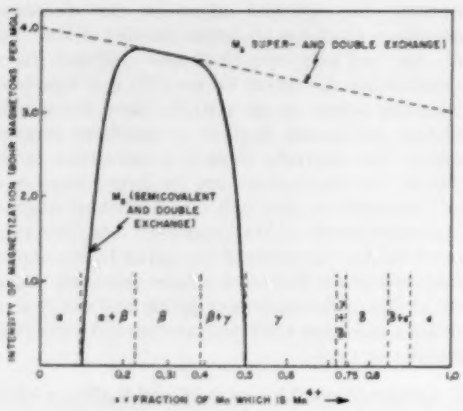


FIG. 6. Predicted intensity of magnetization M_s and phase diagram for the system. $[La_xM(II)_{1-x}]MnO_3$ according to covalent-bond, semicovalent-exchange model.

lie along a Mn−O−Mn axis. This is the optimum configuration for either Mn−O covalent-bond formation or for double exchange. It follows, therefore, that the more stable the tetrahedral (sp^3) orbital associated with the large cation, the higher the temperature at which the lattice bonds order or, in the case of ferromagnetism via double exchange, the higher the Curie temperature.

Jonker and van Santen have found a maximum ferromagnetic Curie temperature at $x\approx0.3$ for Ca-, Sr-, and Ba-substituted manganites; these maximum temperatures are, respectively, 270°K, 375°K, and 350°K. Since the stability of the M^{2+}−O covalent bond increases from Ca^{2+} to Sr^{2+} and Ba^{2+}, the increase of 100°C in the maximum Curie temperature can be understood as due to the differing amounts of covalence associated with the large-cation bonds. Apparently the reduction in Mn−O bond strength with increasing lattice parameter compensates for the increased covalent-bond strength of the large cation on going from $(La,Sr)MnO_3$ to $(La,Ba)MnO_3$.

V. SUMMARY

The predictions of Sec. IV are summarized in Fig. 6 where the compositional regions marked α, $\alpha+\beta$, β, $\beta+\gamma$, γ, $\gamma+\delta$, δ, $\delta+\epsilon$, and ϵ correspond to the following phases:

Phase α:

(i) Orthorhombic ($a_2<a_3=a_1$). At $x=0$ $\Delta\equiv(a_3-a_2)/a_3=0.042$.

(ii) Antiferromagnetic type A with $T_\theta>T_c=T_c^s$ at $x=0$, where T_θ is the covalent-bond ordering temperature, T_c^s is the Curie temperature for semicovalent exchange.

(iii) As x increases, (a_3-a_2) and T_θ decrease while T_c increases; the Mn^{4+} ions couple parallel to the Mn^{3+} ions they replace.

(iv) High resistivity.

Phase $\alpha+\beta$:

(i) Orthorhombic ($a_2<a_3=a_1$) plus cubic.

(ii) Antiferromagnetic type A plus ferromagnetic with $T_\theta>T_c$ and $T_c^s<T_c<T_c^d$, where T_c^d is Curie temperature for double exchange and $T_c^d>T_c^s$ is assumed.

(iii) As x increases, T_θ decreases toward T_c^d, T_c increases toward T_c^d, and resistivity decreases sharply.

Phase β:

(i) Cubic. (Rhombohedral possible).

(ii) Ferromagnetic with $T_\theta=T_c\approx T_c^d$.

(iii) Resistivity low.

(iv) Maximum T_c and minimum resistivity at $x=0.31$.

Phase $\beta+\gamma$:

(i) Cubic plus tetragonal ($c/a<1$).

(ii) Ferromagnetic plus antiferromagnetic type CE with $T_\theta>T_c$ and $T_c^d>T_c>T_c^s$.

(iii) As x increases, T_c decreases toward T_c^s, T_θ and resistivity increase.

Phase γ:

(i) Tetragonal ($c/a<1$) with $(c-a)\approx\Delta/2$ at $x=0.5$.

(ii) Antiferromagnetic type CE with $T_\theta>T_c=T_c^s$ at $x=0.5$.

(iii) As x increases, $(c-a)$ and T_θ decrease, T_c passes through maximum, and the added Mn^{4+} ions couple antiparallel to the Mn^{3+} ions they replace.

(iv) High resistivity.

Phase $\gamma+\delta$:

(i) Tetragonal ($c/a<1$) plus tetragonal ($c/a>1$).

(ii) Antiferromagnetic types CE and C with $T_c>T_c^s$.

(iii) As x increases, T_c decreases and T_θ passes through a minimum value.

(iv) High resistivity.

Phase δ:

(i) Tetragonal ($c/a>1$) with $(c-a)\approx\Delta/2$ at $x=0.75$.

(ii) Antiferromagnetic type C with $T_\theta>T_c>T_c^s$.

(iii) As x increases, $(c-a)$, T_θ, and T_c decrease, and added Mn^{4+} ions couple parallel to the Mn^{3+} ions they replace.

(iv) High resistivity.

Phase $\delta+\epsilon$:

(i) Tetragonal ($c/a>1$) plus cubic.

(ii) Antiferromagnetic types C and G with $T_\theta>T_c>T_c^s$.

(iii) As x increases, T_c decreases and T_θ passes through a minimum.

(iv) High resistivity.

Phase ϵ:

(i) Cubic.

(ii) Antiferromagnetic type G with $T_c\geq T_c^s$.

(iii) As x increases, T_c decreases to T_c^s at $x=1$ and the Mn^{3+} ions couple antiparallel to the Mn^{4+} ions they replace.

(iv) High resistivity.

It should be noted that the $Mn^{4+}-O$ bond has been assumed equal in length to the $Mn^{3+}-O$ bond with covalent character. This is a fair approximation if the covalent character of the $Mn^{4+}-O$ bond is less than that in the $Mn^{3+}-O$ bond; the general decrease in T_0 with added Mn^{4+} content is consistent with this hypothesis.

It is concluded, therefore, that at least in the manganites $[La,M(II)]MnO_3$ the theory of superexchange is inadequate whereas the concepts of covalence and semicovalent exchange offer a consistent model for explaining complicated variations with chemical composition of crystal structure and magnetic interactions. These concepts were first developed to account for cation ordering, lattice distortions, and magnetic interactions in spinel-type lattices[1]; their generality demands that they be valid for other crystal types. It is therefore significant that they apply so strikingly to this perovskite-type system.

ACKNOWLEDGMENTS

I should like to thank W. C. Koehler and E. O. Wollan for their cooperation in notifying me of the progress of their work as it developed and for several helpful suggestions.

APPENDIX

Consider the following problem: assume that when a La^{3+} ion is replaced by a M^{2+} ion, two Mn^{3+} ions become Mn^{4+} and a neighboring La^{3+} becomes La^{2+} so as to minimize the Madelung energy of the configuration. Calculate the fraction of M^{2+} ions necessary to make half the manganese ions Mn^{4+} if the M^{2+} ions replace lanthanum ions randomly and the replacement of a La^{3+} by M^{2+} causes no change in the valencies of the manganese ions. This problem has identical boundary conditions and is mathematically similar to the problem of calculating the fraction of randomly distributed Mn^{4+} ions which gives the largest number of Mn^{3+} ions with one, and only one, Mn^{4+} near neighbor.

Two Mn^{4+} ions are created by the first La^{3+} ion replaced by a M^{2+} ion. When the second lanthanum ion is replaced, the probability that it be a La^{3+} ion is $p_1=1/(N-1)$, and that it be a La^{3+} ion is $(1-p_1)$, where N is the total number of lanthanum ions present. When the third lanthanum ion is replaced, the probability that it be a La^{3+} ion is $p_2=(2-p_1)/(N-2)$, and that it be a La^{3+} ion is $(1-p_2)$. When the $(n+1)$th lanthanum ion is replaced, the probability that it be a La^{3+} ion is $p_n=(n-p_{n-1})/(N-n)$, and that it be a La^{3+} ion is $(1-p_n)$. Half the manganese ions are Mn^{4+} when

$$2\left(n-\sum_n \frac{n-p_{n-1}}{N-n}\right)=\frac{N}{2}. \tag{1}$$

When N is large, the summation can be approximated by

$$\sum_n \frac{n}{N-n}\approx N\int_0^x \frac{x'dx'}{1-x'}=-N[x+\log(1-x)], \tag{2}$$

where $x=n/N$. Equation (1) reduces, therefore, to

$$2x=0.25-\log(1-x),$$
$$x=0.31.$$

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Symmetry Properties of the Energy Bands of the Zinc Blende Structure

R. H. PARMENTER
RCA Laboratories, Radio Corporation of America, Princeton, New Jersey
(Received May 13, 1955)

The symmetry properties of the one-electron energy bands of a crystal of the zinc blende structure are studied by means of group theory. This is done both with and without the inclusion of spin-orbit coupling. The character tables and compatability tables are obtained for the various irreducible representations of the space group T_d^2 associated with the zinc blende structure. The degeneracies and the gradients of the various possible energy bands are studied at lines and points of symmetry in the Brillouin zone. These results are compared with those for the equivalent energy bands in a crystal of the diamond structure.

I. INTRODUCTION

COMPOUNDS of the zinc blende structure are today assuming positions of increasing importance and interest in solid-state physics. The III–V compounds such as InSb are being intensively studied as semiconductors,[1] while II–VI compounds such as ZnS are used both as photoconductors and as phosphors.[2] There is thus much interest in the nature of the one-electron energy bands in these compounds. Considerable information of a qualitative nature can be deduced about the energy-band maxima and minima in these compounds by means of a judicious combination of symmetry arguments, perturbation theory, and experimental information.[3] Here we propose to determine what can be said

[1] H. Welker, Z. Naturforsch. **7a**, 744 (1952); **8a**, 248 (1953).
[2] R. H. Bube, Phys. Rev. **83**, 393 (1951).
[3] F. Herman, J. Electronics (to be published); Phys. Rev. (to be published).

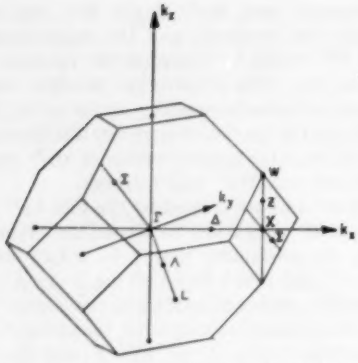


FIG. 1. Lines and points of symmetry in the Brillouin zone of the zinc blende structure

about the energy bands in these compounds on the basis of symmetry alone. In order to do this we will first construct the group-theoretic character tables associated with the zinc blende structure. These character tables are the basic quantities from which all symmetry properties will be deduced. As was first emphasized by Elliott[4] and by Dresselhaus, Kip, and Kittel,[5] the spin-orbit coupling energy,

$$(\hbar/4m^2c^2)[(\nabla V\times\mathbf{p})\cdot\boldsymbol{\sigma}], \tag{1}$$

may have a profound effect upon the one-electron energy levels in semiconductors. (Here V is the crystal potential, $\mathbf{p}$ the electron momentum operator, and $\boldsymbol{\sigma}$ the electron spin operator.) Therefore we will consider the symmetry problem first without and then with the inclusion of spin. Since the diamond structure may be considered to be a special case of the zinc blende structure (i.e., a IV–IV compound), we will compare some of the symmetry properties of the energy bands of the zinc blende structure with those of the diamond structure, character tables for the latter being available both with[6] and without[7] spin.

II. CHARACTER TABLES

The space groups of many of the simpler crystal structures are referred to as *point* space groups.[8] Every element of a point space group may be written as the product of an element of a translation group with an element of a point group, every element of the latter being either a rotation or the product of a rotation and an inversion. For example, the zinc blende structure has the point space group T_d^2, associated with which are the face-centered cubic translation group and the tetrahedral point group T_d. Similarly, the face-centered cubic structure has the point space group O_h^5, associated with which are the face-centered cubic translation group and the full cubic point group O_h. (The space group for the diamond structure, O_h^7, however, is not a point space group.)

An electronic wave function without spin is a scalar function of position and always transforms into itself under a 360° rotation; with spin included, however, the wave function is a spinor and always transforms into the negative of itself under any 360° rotation.[9] Thus, if without spin a crystallographic point group contains n elements C_i, then with spin the point group will contain $2n$ elements, namely n elements C_i and n elements $\bar{C}_i$, where $\bar{C}_i$ is defined as

$$\bar{C}_i\equiv\bar{E}C_i, \tag{2}$$

$\bar{E}$ being defined as some 360° rotation (E is the identity element). C_i and $\bar{C}_i$ are of course identical in their operation on a point in space. The crystallographic groups without and with spin are referred to as single and double groups, respectively.

Since in Schrödinger's equation the Hamiltonian operator H is such that H and H^* describe equivalent physical situations; i.e., H^* differs from H only by some unitary transformation U,

$$H^*=U^{-1}HU, \tag{3}$$

it immediately follows[10] that any eigenfunction of H, namely $\Psi(\mathbf{r},t)$, is degenerate with $U\Psi^*(\mathbf{r},-t)$, also an eigenfunction of H. Without electron spin in the Hamiltonian, $U=1$. With spin, $U=\sigma_y$, the y-component of the Pauli spin operator. This result is referred to as time-reversal symmetry in the Schrödinger equation[6,11,12] in addition to the crystallographic symmetries present. Time-reversal symmetry clearly leads to inversion symmetry of the energy bands in k-space. Since time-reversal symmetry is distinct from the symmetries associated with the space group, it will be necessary to consider the effects of the former after having constructed the character tables for the space group T_d^2.

The shape of the Brillouin zone in k-space associated with the zinc blende structure will be taken identical with that conventionally chosen for the diamond[7] and the face-centered cubic[13] structures, namely a truncated octahedron. Along with the lines and points of symmetry, it is shown in Fig. 1. It should be emphasized, however, that this choice of shape is not unique. In fact, the only parts of the surface of the Brillouin zone which are uniquely specified by symmetry are the lines Z

[4] Private communication to E. N. Adams, II. See E. N. Adams, II, Phys. Rev. **92**, 1063 (1953), reference 7.

[5] Dresselhaus, Kip, and Kittel, Phys. Rev. **95**, 568 (1954).

[6] R. J. Elliott, Phys. Rev. **96**, 280 (1954). Elliott's character table for the extra representations of Λ and F is in error. The correct table is identical with Table III of the present paper; i.e., the character table for Λ and F is identical for the simple cubic, body-centered cubic, face-centered cubic, diamond, and zinc blende structures.

[7] C. Herring, J. Franklin Inst. **233**, 525 (1942). Table XI of this reference is in error, but it is corrected in Table VIII of reference 6.

[8] W. H. Zachariasen, *Theory of X-Ray Diffraction in Crystals* (John Wiley and Sons, Inc. New York, 1945), Chap. 2.

[9] See, for example, E. Feenberg and G. E. Pake, *Notes on the Quantum Theory of Angular Momentum* (Addison-Wesley Press, Cambridge, 1953), p. 28.

[10] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (John Wiley and Sons, Inc., New York, 1952), p. 525.

[11] E. Wigner, Gött. Nachr. p. 546 (1932).

[12] C. Herring, Phys. Rev. **52**, 361 (1937).

[13] Bouckaert, Smoluchowski, and Wigner, Phys. Rev. **50**, 58 (1936).

and the points L lying in the centers of the hexagonal faces. As far as the space group T_d^2 is concerned, L has no higher symmetry than that of Λ, so that L will not be mentioned in constructing the character tables. The additional symmetry of L arises from time-reversal.

Consider a given point space group. For a given wave vector $\mathbf{k}$ associated with some symmetry point in k-space, we find those operations of the point group associated with the space group which transform $\mathbf{k}$ into itself plus some reciprocal-lattice vector. Such a set of operations forms a group called the group of the wave vector $\mathbf{k}$. There will be a distinct wave-vector group for each type of symmetry line and point in the Brillouin zone. All the wave-vector groups are subgroups of the point group associated with the space group. We now determine the members of the wave-vector group for zinc blende for each type of symmetry point in the Brillouin zone and compare each of these groups with the corresponding wave-vector group for the face-centered cubic structure as given in reference 13 (without spin) and reference 6 (with spin). (We use the terminology of reference 13 to denote the group operations and the lines and points of symmetry.) For Λ, the wave-vector groups for zinc blende and fcc are identical, and thus so are the character tables (note comment of reference 6). For both Γ and X. the wave-vector group for fcc is the direct product of the corresponding wave-vector group for zinc blende and the group of order two, $\mathfrak{C}_2$, whose elements are E, the identity element, and J, the inversion element.[14] Group theory states that, considered as a matrix,[15] the character table appropriate to a direct-product group is itself the direct product[16] of the matrices representing the character tables of the groups occurring as factors of the direct-product group.[17] With the aid of the character table of the group $\mathfrak{C}_2$ (it is equivalent to Table IV), it is a simple matter to obtain the character tables for zinc blende for the points Γ and X from the corresponding character tables for fcc. Considering the Σ and Z lines in zinc blende, it is apparent that without spin these wave-vector groups each contain only two elements and are thus both simply isomorphic with the group $\mathfrak{C}_2$. Including spin, it is evident that these two groups are simply isomorphic with the cyclic group of order four. Thus for each group the characters are the appropriate powers of a fourth root of unity.[18] Turning to the point W, without spin the wave-vector group is also the fourth-order cyclic group, the elements of the group being various powers of JC_4, an inversion combined with a 90° rotation about the fourfold axis parallel to the face diagonal containing W.

TABLE I. Character table for the representations of the single group of Γ.

24	Γ	Γ_1	Γ_2	Γ_{12}	Γ_{15}	Γ_{25}
1	E	1	1	2	3	3
3	C_4^2	1	1	2	−1	−1
8	C_3	1	1	−1	0	0
6	JC_4	1	−1	0	−1	1
6	JC_2	1	−1	0	1	−1

TABLE II. Character table for the representations of the single group of Δ. (The two distinct operations JC_2 are those about the two twofold axes perpendicular to Δ.) Δ_3 and Δ_4 stick together.

4	Δ	Δ_1	Δ_2	Δ_3	Δ_4
1	E	1	1	1	1
1	JC_2	1	−1	1	−1
1	JC_2	1	−1	−1	1
1	$C_4^2\|\|$	1	1	−1	−1

TABLE III. Character table for the representations of the single group of Λ.

6	Λ	Λ_1	Λ_2	Λ_3
1	E	1	1	2
2	C_3	1	1	−1
3	JC_2	1	−1	0

TABLE IV. Character table for the representations of the single groups of Σ and Z.

2	Σ	Z	Σ_1	Σ_2
1	E	E	1	1
1	JC_2	$C_4^2\|\|$	1	−1

TABLE V. Character table for the representations of the single group of X.

8	X	X_1	X_2	X_3	X_4	X_5
1	E	1	1	1	1	2
2	$C_4^2\|\|$	1	1	−1	−1	0
1	$C_4^2\|\|$	1	1	1	1	−2
2	$JC_{4\|\|}$	1	−1	−1	1	0
2	JC_2	1	−1	1	−1	0

TABLE VI. Character table for the representations of the single group of W. (The two distinct operations JC_4 are those about the fourfold axis parallel to the face diagonal containing W.) W_3 and W_4 are degenerate by time-reversal symmetry.

4	W	W_1	W_2	W_3	W_4
1	E	1	1	1	1
1	JC_4	1	−1	i	$-i$
1	JC_4	1	−1	$-i$	i
1	C_4^2	1	1	−1	−1

[14] Consider a group $\mathfrak{A}$ with elements a_i and a group $\mathfrak{B}$ with elements b_j, where $a_ib_j=b_ja_i$. The direct product of groups $\mathfrak{A}$ and $\mathfrak{B}$ is a group $\mathfrak{C}=\mathfrak{A}\times\mathfrak{B}$ whose elements are $c_{ij}=a_ib_j$.

[15] A character table may be considered a square matrix whose rows designate the classes of elements and whose columns designate the irreducible representations.

[16] The direct product of an $m\times m$ matrix having elements α_{ij} and an $n\times n$ matrix having elements $\beta_{i'j'}$ is an $mn\times mn$ matrix having elements $\alpha_{ij}\beta_{i'j'}$, i and i' designating the row and j and j' the column.

[17] F. D. Murnaghan, *The Theory of Group Representations* (Johns Hopkins Press, Baltimore, 1938), p. 101.

[18] The cyclic group of order n contains the elements C, C^2, C^3, $\cdots C^n=E$. Since any two elements commute, each element forms a class by itself. Since there are n classes, there are n representations, each necessarily one-dimensional. A one-dimensional representation and its character being identical, the characters must transform as do their corresponding group elements and thus must be appropriate powers of an nth root of unity.

TABLE VII. Character table for the extra representations of the double group of Γ.

48	Γ	Γ_6	Γ_7	Γ_8
1	E	2	2	4
1	$\bar{E}$	−2	−2	−4
6	C_4^2, $\bar{C}_4^2$	0	0	0
8	C_3	1	1	−1
8	$\bar{C}_3$	−1	−1	1
6	JC_4	$\sqrt{2}$	$-\sqrt{2}$	0
6	$J\bar{C}_4$	$-\sqrt{2}$	$\sqrt{2}$	0
12	JC_2, $J\bar{C}_2$	0	0	0

TABLE VIII. Character table for the extra representation of the double group of Δ.

8	Δ	Δ_6
1	E	2
1	$\bar{E}$	−2
2	JC_2, $J\bar{C}_2$	0
2	JC_2, $J\bar{C}_2$	0
2	$C_4^2\|\|$, $\bar{C}_4^2\|\|$	0

TABLE IX. Character table for the extra representations of the double group of Λ, Λ_4, and Λ_5 are degenerate by time-reversal symmetry.

12	Λ	Λ_4	Λ_5	Λ_6
1	E	1	1	2
1	$\bar{E}$	−1	−1	−2
2	C_3	−1	−1	1
2	$\bar{C}_3$	1	1	−1
3	JC_2	i	$-i$	0
3	$J\bar{C}_2$	$-i$	i	0

Similarly, with spin the wave-vector group is the eighth-order cyclic group, so that the characters are the appropriate powers of an eighth root of unity. For the fourfold axis Δ, without spin the wave-vector group contains the identity element E, the two commuting elements JC_2 (representing reflections in the planes perpendicular to the twofold axes which are in turn perpendicular to Δ), and the element $C_4^2\|\|$ (representing a 180° rotation about Δ) which is the product of the two reflections. Thus the group may be considered to be the direct product of two second-order groups $\mathfrak{C}_2$, and the character table follows immediately. With spin the wave-vector group for Δ can be seen to be simply isomorphic with the group for X without spin, with the result that the two character tables are equivalent.

In this simple fashion we have obtained the character tables for the zinc blende structure for both the single and the double groups. The single groups without spin are listed in Tables I–VI. A perusal of our results demonstrates a rule generally valid[19] that the irreducible representations of the single group occur unchanged in the double group with the elements C_i and $\bar{C}_i$ having the same character. For this reason, we do not bother to list these representations again in the double groups, but restrict ourselves to listing the additional or *extra* representations, which are given in Tables VII–XII. In contrast with the other representations, the extra representations always occur with the elements C_i and $\bar{C}_i$ having characters of opposite sign. (Thus whenever C_i and $\bar{C}_i$ occur in the same class, the corresponding characters of the extra representations always vanish.)

[19] W. Opechowski, Physica 7, 552 (1940).

III. RESULTS

As we move continuously in k-space from a point of higher symmetry to a point of lower symmetry, we wish to know how the various irreducible representations associated with the former point transform into those associated with the latter point. To do this we construct what are known as compatability tables.[13] We express the characters of a representation for the higher-symmetry point in terms of a sum of corresponding characters of lower-symmetry representations, this to be done simultaneously for each group operation common to the two points. These lower-symmetry representations are what combine at the higher-symmetry point to form the higher-symmetry representation. The compatability tables for the single groups without spin are given in Table XIII, and those for the double groups with spin (extra representations) are given in Table XIV.

Physically, the zinc blende structure may be considered to be the result of a deformation upon either the face-centered cubic structure, O_h^5, or the diamond structure O_h^7. Thus we may construct a second type of

TABLE X. Character table for the extra representations of the double groups of Σ and Z. Σ_3 and Σ_4, Z_3 and Z_4, are degenerate by time-reversal symmetry.

4	Σ	Z	Σ_3	Σ_4
1	E	E	1	1
1	$\bar{E}$	$\bar{E}$	−1	−1
1	JC_2	$C_4^2\perp$	i	$-i$
1	$J\bar{C}_2$	$\bar{C}_4^2\perp$	$-i$	i

TABLE XI. Character table for the extra representations of the double group of X.

16	X	X_6	X_7
1	E	2	2
1	$\bar{E}$	−2	−2
4	$C_4^2\perp$, $\bar{C}_4^2\perp$	0	0
2	$C_4^2\|\|$, $\bar{C}_4^2\|\|$	0	0
2	$JC_4\|\|$	$\sqrt{2}$	$-\sqrt{2}$
2	$J\bar{C}_4\|\|$	$-\sqrt{2}$	$\sqrt{2}$
4	JC_2, $J\bar{C}_2$	0	0

TABLE XII. Character table for the extra representations of the double group of W. W_5 is degenerate with W_7 and W_6 is degenerate with W_8 by time-reversal symmetry. $\epsilon=(1+i)/\sqrt{2}$.

8	W	W_5	W_6	W_7	W_8
1	E	1	1	1	1
1	$\bar{E}$	−1	−1	−1	−1
1	JC_4	ϵ	$-\epsilon$	ϵ^*	$-\epsilon^*$
1	$J\bar{C}_4$	$-\epsilon$	ϵ	$-\epsilon^*$	ϵ^*
1	JC_4	ϵ^*	$-\epsilon^*$	ϵ	$-\epsilon$
1	$J\bar{C}_4$	$-\epsilon^*$	ϵ^*	$-\epsilon$	ϵ
1	C_4^2	i	i	$-i$	$-i$
1	$\bar{C}_4^2$	$-i$	$-i$	i	i

TABLE XIII. Compatability tables for the representations of the single groups connecting symmetry points with symmetry axes.

Γ	Δ	Λ	Σ
Γ_1	Δ_1	Λ_1	Σ_1
Γ_2	Δ_2	Λ_2	Σ_2
Γ_{12}	$\Delta_1+\Delta_2$	Λ_3	$\Sigma_1+\Sigma_2$
Γ_{15}	$\Delta_1+\Delta_2+\Delta_4$	$\Lambda_1+\Lambda_3$	$2\Sigma_1+\Sigma_2$
Γ_{25}	$\Delta_2+\Delta_2+\Delta_4$	$\Lambda_2+\Lambda_3$	$\Sigma_1+2\Sigma_2$
X	Δ	Σ	Z
X_1	Δ_1	Σ_1	Z_1
X_2	Δ_2	Σ_2	Z_1
X_3	Δ_1	Σ_1	Z_2
X_4	Δ_2	Σ_2	Z_2
X_5	$\Delta_2+\Delta_4$	$\Sigma_1+\Sigma_2$	Z_1+Z_2
W	Z		
W_1	Z_1		
W_2	Z_1		
W_3	Z_2		
W_4	Z_2		

TABLE XIV. Compatability tables for the extra representations of the double groups connecting symmetry points with symmetry axes.

Γ	Δ	Λ	Σ
Γ_6	Δ_5	Λ_6	$\Sigma_3+\Sigma_4$
Γ_7	Δ_5	Λ_6	$\Sigma_3+\Sigma_4$
Γ_8	$2\Delta_5$	$\Lambda_4+\Lambda_5+\Lambda_6$	$2\Sigma_3+2\Sigma_4$
X	Δ	Σ	Z
X_6	Δ_5	$\Sigma_3+\Sigma_4$	Z_3+Z_4
X_7	Δ_5	$\Sigma_3+\Sigma_4$	Z_3+Z_4
W	Z		
W_5	Z_3		
W_6	Z_3		
W_7	Z_4		
W_8	Z_4		

compatability table between T_d^2 and O_h^5 and between T_d^2 and O_h^7, showing how the various irreducible representations for O_h^5 and O_h^7 are transformed by the deformations which convert these structures into T_d^2. Using the same technique described in the previous paragraph, we obtain Tables XV and XVI. Such tables are particularly useful in looking at some zinc blende structure as a perturbation upon some similar diamond structure,[3] e.g., GaAs as a perturbation upon Ge.

A third type of compatibility table is needed to show how the various irreducible representations for zinc blende without spin are transformed upon the introduction of spin. Such information is given in Table XVII, which demonstrates the general rule that only the *extra* irreducible representations occur.[6] Table XVII is obtained in the following fashion. If we take the point group T_d and replace the elements JC_4 (90° rotation about a fourfold axis followed by an inversion) and JC_2 (180° rotation about a twofold axis followed by an inversion) by C_4 and C_2 respectively, we obtain the point group O, which can be seen to be simply isomorphic with T_d. Since O is a subgroup of the rotation group δ_3, it follows that a conventional compatability table can be set up connecting the representations of δ_3 with those of T_d. A wave function of spin alone belongs to the $D_{\frac{1}{2}}$ representation of δ_3. Since the explicit form of the $D_{\frac{1}{2}}$ representation matrices is available in the literature[6,19] the characters can be calculated, whence it is found that $D_{\frac{1}{2}}$ corresponds to the Γ_6 representation of the T_d point group. An eigenfunction of our Hamiltonian containing spin for the zinc blende structure will be a sum of products of a function of spin (belonging to the $D_{\frac{1}{2}}$ representation) and a function of position (belonging to some representation Ξ_i of the single space group T_d^2). Thus this eigenfunction will belong to one of the irreducible representations contained in the direct product $\Xi_i \times \Gamma_6$. Table XVII is now obtained by making use of the fact that a character of the direct product of two representations, which is the product of the corresponding characters of the two representations, can be written as the sum of the corresponding characters of all the irreducible representations composing the direct product.

We are now in a position to study the nature of the energy bands in zinc blende (as imposed by symmetry). First we consider the possibility of the sticking together of two bands associated with distinct representations; i.e., two wave functions of a given $\mathbf{k}$ being degenerate although belonging to two distinct representations. An operation of the class $JC_4^{(1)}$ (rotation by 90° about the axis Δ followed by an inversion) followed by the time-reversal symmetry operation (replacing $\mathbf{k}$ by $-\mathbf{k}$) will convert a wave function belonging to Δ_3 into one

TABLE XV. Compatability tables for the representations of the single groups connecting the zinc blende (T_d^2) with the face-centered-cubic (O_h^5) and the diamond (O_h^7) structures.

T_d^2	O_h^5	T_d^2	O_h^7
Γ_1	Γ_1 or $\Gamma_{2'}$	Γ_1	Γ_1 or $\Gamma_{2'}$
Γ_2	Γ_2 or $\Gamma_{1'}$	Γ_2	Γ_2 or $\Gamma_{1'}$
Γ_{12}	Γ_{12} or $\Gamma_{12'}$	Γ_{12}	Γ_{12} or $\Gamma_{12'}$
Γ_{15}	Γ_{15} or $\Gamma_{25'}$	Γ_{15}	Γ_{15} or $\Gamma_{25'}$
Γ_{25}	Γ_{25} or $\Gamma_{15'}$	Γ_{25}	Γ_{25} or $\Gamma_{15'}$
Δ_1	Δ_1 or $\Delta_{2'}$	Δ_1	Δ_1 or $\Delta_{2'}$
Δ_2	Δ_2 or $\Delta_{1'}$	Δ_2	Δ_2 or $\Delta_{1'}$
Δ_2, Δ_4	Δ_5	Δ_2, Δ_4	Δ_5
Λ_1	Λ_1	Λ_1	Λ_1
Λ_2	Λ_2	Λ_2	Λ_2
Λ_3	Λ_3	Λ_3	Λ_3
Σ_1	Σ_1 or Σ_3	Σ_1	Σ_1 or Σ_3
Σ_2	Σ_2 or Σ_4	Σ_2	Σ_2 or Σ_4
Z_1	Z_1 or Z_2	Z_1, Z_2	Z_1
Z_2	Z_3 or Z_4		
X_1	X_1 or $X_{2'}$	X_1, X_3	X_1
X_2	X_2 or $X_{1'}$		
X_3	X_3 or $X_{4'}$	X_2, X_4	X_2
X_4	X_4 or $X_{3'}$		
X_5	X_5 or $X_{5'}$	X_5	X_3 or X_4
W_1	W_1 or W_2	W_1, W_3	W_1
W_2	$W_{1'}$ or $W_{2'}$		
W_3, W_4	W_3	W_2, W_4	W_2

TABLE XVI. Compatability tables for the extra representations of the double groups connecting the zinc blende (T_d^2) with the face-centered cubic (O_h^5) and the diamond (O_h^7) structures.

T_d^2	O_h^5	T_d^2	O_h^7
Γ_6	Γ_6^+ or Γ_7^-	Γ_6	Γ_6^+ or Γ_7^-
Γ_7	Γ_6^- or Γ_7^+	Γ_7	Γ_6^- or Γ_7^+
Γ_8	Γ_8^+ or Γ_8^-	Γ_8	Γ_8^+ or Γ_8^-
Δ_5	Δ_5 or Δ_7	Δ_5	Δ_5 or Δ_7
Λ_4	Λ_4	Λ_4	Λ_4
Λ_5	Λ_5	Λ_5	Λ_5
Λ_6	Λ_6	Λ_6	Λ_6
Σ_3, Σ_4	Σ_3	Σ_3, Σ_4	Σ_3
Z_3, Z_4	Z_3	Z_3, Z_4	Z_3
X_6	X_6^+ or X_7^-	X_6, X_7	X_6
X_7	X_6^- or X_7^+		
W_5, W_7	W_5	W_5	W_5 or W_6 or W_7
W_6, W_8	W_7	W_6	W_4 or W_8 or W_7
		W_7	W_3 or W_5 or W_7
		W_8	W_4 or W_6 or W_7

belonging to Δ_4, while leaving $\mathbf{k}$ and the Hamiltonian invariant. Thus Δ_3 and Δ_4 stick together. In a similar fashion, at the point L the time-reversal symmetry operation will convert a wave function belonging to Λ_4 into Λ_5, while leaving $\mathbf{k}$ and the Hamiltonian invariant. Thus Λ_4 and Λ_5 stick together at the point L. These are the only cases of bands sticking together as a result of time-reversal symmetry. Since, however, the characters belonging to the representation Z_3 are complex conjugates of those belonging to Z_4, it follows from time-reversal symmetry that for every wave function of a given wave vector $\mathbf{k}$ belonging to Z_3 there is a degenerate wave function of $-\mathbf{k}$ belonging to Z_4. Identical arguments hold for Λ_4 and Λ_5, for Σ_3 and Σ_4, for W_3 and W_4, W_5 and W_7, and W_6 and W_8, respectively.[20]

As was emphasized by Elliott,[21] at a point of no special symmetry in k-space, the extra representations are always doubly degenerate in crystals possessing a center of inversion, such as diamond. Because of the lack of inversion symmetry, this is not the case in zinc blende.

We wish to investigate whether or not, on the basis of symmetry, the gradient in k-space of a given energy band will vanish in a given direction at a given point in k-space. Since $\nabla_k E(\mathbf{k})$ is proportional to the expectation value of momentum, it is necessary to investigate whether or not the matrix elements of $(\hbar/i)\nabla$ vanish, these matrix elements being taken with respect to the wave functions belonging to a particular irreducible representation of the wave-vector group associated with the given point in k-space. For the component of the gradient along a direction specified by the unit vector $\mathbf{u}$, we need to consider matrix elements of $(\hbar/i)\mathbf{u}\cdot\nabla$, these matrix elements being integrals over ordinary space.[22] The integrands of these integrals transform into one another under those particular operations associated with the given wave vector group which transform $\mathbf{u}\cdot\nabla$ into either plus or minus itself (the operation acting on ∇ but not on $\mathbf{u}$). Such operations form a group which we shall call the gradient wave-vector group. If we knew the character tables of the gradient wave-vector groups, it would be an easy matter to decide whether or not a given matrix element vanishes. Thus we would resolve the representation to which the integrand belongs (the direct product of the three representations corresponding to the two wave functions in the integrand and to $\mathbf{u}\cdot\nabla$, respectively) into a sum of irreducible representations of the gradient wave vector group, in exactly the same manner used to obtain Table XVII. We now make use of the fact that the integral will vanish unless the sum of irreducible representations contains the totally symmetric representation (that representation all of whose characters are unity).[23,24] We actually have the character tables of the gradient wave-vector groups, since every gradient wave-vector group is simply isomorphic with some particular wave-vector group.[25]

TABLE XVII. Compatability table connecting the representations of the single groups with the extra representations of the double groups.

Ξ_i	$\Xi_i \times \Gamma_6$	Ξ_i	$\Xi_i \times \Gamma_6$
Γ_1	Γ_6	Σ_1	$\Sigma_3+\Sigma_4$
Γ_2	Γ_7	Σ_2	$\Sigma_3+\Sigma_4$
Γ_{12}	Γ_8		
Γ_{15}	$\Gamma_7+\Gamma_8$	Z_1	Z_3+Z_4
Γ_{25}	$\Gamma_6+\Gamma_8$	Z_2	Z_3+Z_4
Δ_1	Δ_5	X_1	X_6
Δ_2	Δ_5	X_2	X_7
Δ_3	Δ_5	X_3	X_7
Δ_4	Δ_5	X_4	X_6
		X_5	X_6+X_7
Λ_1	Λ_6		
Λ_2	Λ_6	W_1	W_5+W_7
Λ_3	$\Lambda_4+\Lambda_5+\Lambda_6$	W_2	W_5+W_8
		W_3	W_5+W_8
		W_4	W_6+W_7

[20] The writer is indebted to G. Dresselhaus, who has independently studied the symmetry properties of the zinc blende structure, for pointing out errors in the writer's original analysis of the effect of time-reversal symmetry on the points Λ and Σ for the double groups.

[21] R. J. Elliott, Phys. Rev. 96, 266 (1954).

[22] For the spin case, ∇ should be replaced by $[\nabla+(i/4mc^2)(\boldsymbol{\sigma}\times\nabla V]$. The latter, however, has the same transformation properties as ∇ and need not be considered further.

[23] See, for example, Eyring, Walter, and Kimball, *Quantum Chemistry* (John Wiley and Sons, Inc., New York, 1944), p. 187.

[24] In some cases where the above analysis does not indicate a vanishing gradient, it is evident that the gradient must vanish in order to maintain time-reversal symmetry.

[25] The gradient wave-vector groups at Γ for $\mathbf{u}$ pointing along Δ, Λ, and Σ are simply isomorphic with the wave-vector groups at X, Λ, and Δ, respectively. The gradient wave-vector groups at X for $\mathbf{u}$ pointing along Δ, Σ, and Z are simply isomorphic with the wave-vector groups at X, Δ, and Δ, respectively. The gradient wave-vector groups at W for $\mathbf{u}$ pointing along Z and along either of the two (100) directions perpendicular to Z are simply isomorphic with the wave-vector groups at W and Σ, respectively. The gradient wave-vector groups at Δ for $\mathbf{u}$ pointing along a perpendicular (100) direction and along a perpendicular (110) direction (both perpendicular to Δ) are simply isomorphic with the wave-vector groups at Σ and Δ, respectively. The gradient wave-vector groups at Λ, Σ, and Z are all simply isomorphic with the wave-vector group at Σ.

Finally, we wish to point out two facts which are useful in any study of energy bands. The first is that the complete set of energy bands must have the full symmetry of the Brillouin zone. The second is that, as we move along a continuous curve (with continuous derivatives) in k-space, it is always possible to trace out a continuous curve (with continuous derivatives) of allowed energy *versus* $\mathbf{k}$.

With the aid of the considerations of the previous paragraphs it is now possible to deduce the nature of the energy bands in the vicinity of a symmetry point. First we consider the single groups without spin. For this case the energy bands are identical in their symmetry properties with those of the diamond structure everywhere except along the symmetry axes Σ and Z. As in diamond, Γ is a point of zero slope in the directions of the three symmetry axes Δ, Λ, and Σ; L is a point of zero slope in the direction of Λ; Δ is a point of zero slope along the (100) and (110) directions perpendicular to the given Δ axis; and Λ may be a point of either zero slope or finite slope in directions perpendicular to the given Λ axis (Λ_1 and Λ_2 have zero slope, Λ_3 has finite slope). Although Σ is nondegenerate as in diamond, it differs from diamond in that it is a point of finite slope along the (100) direction perpendicular to the given Σ axis (although it is a point of zero slope along the (110) direction perpendicular to the given Σ axis). If we consider the square-face diagonals, unlike diamond X is always a point of zero slope along the three symmetry axes Δ, Σ, and Z, while Z and W are nondegenerate, W being a point of zero slope in any direction and Z being a point of zero slope in directions normal to the given Z-axis.

We now consider the double groups with spin. For this case the energy bands of the zinc blende structure are considerably different from those of the diamond structure. The types of possible E *versus* $\mathbf{k}$ curves in the neighborhoods of symmetry points are shown in Fig. 2. Γ_6 and Γ_7 are points of zero slope along Δ and Λ but finite slope along Σ, while Γ_8 may be a point of finite slope in any direction (although Λ_6 will approach Γ_8 with zero slope), as has been suggested by Kittel and Dresselhaus.[26] Λ_4 and Λ_5 approach L with finite slope while Λ_6 approaches L with zero slope. X is a point of zero slope along Δ and finite slope along Σ and Z. W is a point of zero slope in any direction. Δ is a point of finite slope in directions normal to the given Δ axis. Λ may be a point of either zero slope or finite slope in directions perpendicular to the given Λ axis (Λ_4 and Λ_5 have zero slope, Λ_6 has finite slope). At Σ or Z the slopes in directions perpendicular to their respective axes are similar to those for Σ and Z in the absence of spin. Unlike diamond, at many points in k-space a nondegenerate band may occur. We see that the only points of special symmetry at which an energy band minimum or maximum can occur on the basis of symmetry alone are the points W. A minimum or maximum *may* occur somewhere along Λ or Z, but none will occur along Δ or Σ, or at the points Γ, X, or L.

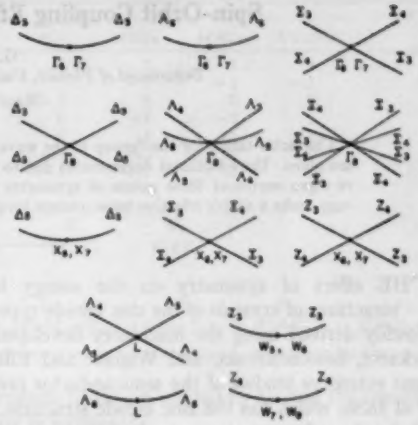


FIG. 2. Nature of possible E *versus* $\mathbf{k}$ curves in the neighborhood of symmetry points.

In conclusion, we might point out that if we look at the zinc blende structure as a perturbation on the diamond structure,[3] then the difference between a given energy level of the unperturbed diamond structure and the corresponding energy level of the perturbed zinc blende structure will vanish (to the accuracy of first-order perturbation theory) everywhere except on the square-face diagonals in the case of no spin, but will vanish only along the axis Δ (including Γ but excluding X) in the case of spin. To prove this, we first note that the perturbation potential which converts a diamond structure into a zinc blende structure will have odd parity with respect to an inversion about a point midway between two adjacent atoms. By combining degenerate unperturbed diamond-structure wave functions associated with different values of $\mathbf{k}$, we can always form new unperturbed wave functions which have either even or odd parity with respect to inversion about such a point midway between two adjacent atoms. It now follows that all diagonal matrix elements of the perturbation potential with respect to these new unperturbed wave functions will vanish, and thus the sum of diagonal matrix elements over a degenerate set of unperturbed wave functions, i.e., the trace, will vanish. This means that first-order perturbation theory may split a degeneracy but will not affect its "center-of-gravity" (mean energy). But Tables XV and XVI show that, in the case of no spin, degeneracies are split only along the square-face diagonals, whereas, when spin is present, degeneracies are split everywhere except along the axis Δ (including Γ but excluding X). This completes the proof.

IV. ACKNOWLEDGMENTS

The writer is greatly indebted to Dr. Frank Herman for many stimulating and enlightening discussions during the course of this investigation.

[26] C. Kittel (private communication).

Spin-Orbit Coupling Effects in Zinc Blende Structures*

G. DRESSELHAUS†
Department of Physics, University of California, Berkeley, California
(Received June 30, 1955)

Character tables for the "group of the wave vector" at certain points of symmetry in the Brillouin zone are given. The additional degeneracies due to time reversal symmetry are indicated. The form of energy vs wave vector at these points of symmetry is derived. A possible reason for the complications which may make a simple effective mass concept invalid for some crystals of this type structure will be presented.

THE effect of symmetry on the energy band structures of crystals of the zinc blende type can be readily derived using the machinery developed by Bouckaert, Smoluchowski, and Wigner[1] and Elliott.[2] Recent extensive studies of the semiconductor properties[3] of InSb, which has the zinc blende structure, and preliminary cyclotron resonance investigations[4] have indicated a need for a more thorough understanding of the possible energy band structures of a zinc blende type crystal.

A zinc blende structure consists of two interpenetrating face centered cubic lattices; each f.c.c. may be considered a sublattice. The two sublattices are displaced by one quarter of a body diagonal and each consists entirely of one species of atom. If the two sublattices are identical, one has a diamond structure. The symmetry properties of diamond are fully discussed in reference 2.

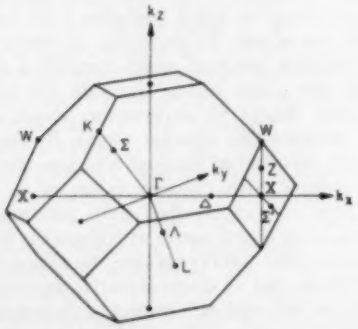


FIG. 1. The first Brillouin zone for a face centered cubic, diamond, and zinc blende structure. Points and lines of symmetry are indicated using the notation of reference 1.

The zinc blende structure has the space group symmetry $F\bar{4}3m$ or T_d^2. There are no glide planes or screw axes, so the group of any wave vector $\mathbf{k}$ has only simple operations. The first Brillouin zone is the well-known truncated octahedron shown in Fig. 1. The character tables for the group of the wave vector $\mathbf{k}$ for certain points of symmetry in the Brillouin zone are given in Tables I through VI. When the spin is included in the problem only the double representations occur (i.e., representations for which a 360° rotation, $\bar{E}$, changes the sign of the wave function). The effect of including spin in the problem is to form wave functions of a spatial function times a spin function which will transform[5] as $D_{\frac{1}{2}}$. The total wave function will then transform as the direct product of a single group representation with $D_{\frac{1}{2}}$. This direct product then can be decomposed in terms of representations of the double group. If more than one representation of the double group occurs in the decomposition of the direct product, a spin-orbit splitting of the level is indicated. A table of the direct products of the single group representations with $D_{\frac{1}{2}}$ is included with each character table. The compatibility relations for certain lines of symmetry are given in Table VII. These relations give the splitting of the degeneracies as one proceeds along the symmetry axes. The extra degeneracies due to time reversal symmetry can be found using the standard test due to Herring[6] and Eliott.[2] The extra degeneracies are indicated in each table.

The principal difference from the diamond structure is the lack of inversion symmetry for the point groups in the zinc blende structure. Without inversion symmetry one still has the result from Kramers' theorem[7] that $E(\mathbf{k})=E(-\mathbf{k})$, but now the periodic part of the Bloch functions no longer satisfies the condition $u_{-\mathbf{k}}(\mathbf{r})=u_{\mathbf{k}}(-\mathbf{r})$, and hence a twofold degeneracy throughout the Brillouin zone is not required.

The one electron Schrödinger equation for the

* This work has been supported in part by the Office of Naval Research and the U. S. Signal Corps.
† Now at the Institute for the Study of Metals, University of Chicago, Chicago, Illinois.
[1] Bouckaert, Smoluchowski, and Wigner, Phys. Rev. **50**, 58 (1936).
[2] R. J. Elliott, Phys. Rev. **96**, 280 (1954).
[3] H. Welker, Z. Naturforsch. **7a**, 744 (1952); **8a**, 248 (1953); M. Tanenbaum and H. B. Briggs, Phys. Rev. **91**, 1561 (1953); G. L. Pearson and M. Tanenbaum, Phys. Rev. **90**, 153 (1953); M. Tanenbaum and J. P. Maita, Phys. Rev. **91**, 1009 (1953); H. Weiss, Z. Naturforsch. **8a**, 463 (1953); O. Madelung and H. Weiss, Z. Naturforsch. **9a**, 527 (1954).
[4] Dresselhaus, Kip, Kittel, and Wagoner, Phys. Rev. **98**, 556 (1955).
[5] E. Wigner, *Gruppentheorie* (J. W. Edwards, Michigan, 1944), p. 245.
[6] C. Herring, Phys. Rev. **52**, 361 (1937).
[7] This theorem states that in the absence of magnetic fields $\Psi_{\mathbf{k}}$ and $i\sigma_y\Psi_{\mathbf{k}}^*$ are solutions of the Hamiltonian for the same energy. The second solution belongs to wave vector $-\mathbf{k}$, and hence we have two solutions at $\mathbf{k}$ and $-\mathbf{k}$ with the same energy.

TABLE I. Character table of the double group of Γ; $\mathbf{k}=(000)$.

Γ	E	$\bar{E}$	$6C_4^2$	$8C_3$	$8\bar{C}_3$	$6I\times C_2$	$6I\times\bar{C}_2$	$12I\times C_2$
Γ_1	1	1	1	1	1	1	1	1
Γ_2	1	1	1	1	1	−1	−1	−1
Γ_3	2	2	2	−1	−1	0	0	0
$\Gamma_4(x,y,z)$	3	3	−1	0	0	−1	−1	1
Γ_5	3	3	−1	0	0	1	1	−1
Γ_6	2	−2	0	1	−1	$\sqrt{2}$	$-\sqrt{2}$	0
Γ_7	2	−2	0	1	−1	$-\sqrt{2}$	$\sqrt{2}$	0
Γ_8	4	−4	0	−1	1	0	0	0

Γ_i	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Γ_6	Γ_7	Γ_8
$\Gamma_i\times D_{\frac{1}{2}}$			Γ_6	Γ_7	Γ_8	$\Gamma_7+\Gamma_8$	$\Gamma_6+\Gamma_8$	$\Gamma_6+\Gamma_7+2\Gamma_8$

Selection rules

Γ_i	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Γ_6	Γ_7	Γ_8
$\Gamma_i\times\Gamma_4$	Γ_4	Γ_5	$\Gamma_4+\Gamma_5$	$\Gamma_1+\Gamma_2+\Gamma_4+\Gamma_5$	$\Gamma_3+\Gamma_3+\Gamma_4+\Gamma_5$	$\Gamma_7+\Gamma_8$	$\Gamma_6+\Gamma_8$	$\Gamma_6+\Gamma_7+2\Gamma_8$

problem with spin-orbit coupling is

$$\left[\frac{p^2}{2m}+V+\frac{\hbar}{4m^2c^2}(\nabla V\times\mathbf{p})\cdot\boldsymbol{\sigma}\right]\Psi_{\mathbf{k}}=E_{\mathbf{k}}\Psi_{\mathbf{k}}. \quad (1)$$

The translational symmetry of the lattice requires that the wave functions be of the Bloch form, i.e.,

$$\Psi_{\mathbf{k}}=u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}},$$

where $u(\mathbf{r})$ is periodic and satisfies the equation

$$\left[\frac{\mathbf{p}^2}{2m}+V+\frac{\hbar}{4m^2c^2}(\nabla V\times\mathbf{p})\cdot\boldsymbol{\sigma}\right]u_{\mathbf{k}}$$

$$+\hbar\mathbf{k}\cdot\left(\frac{\mathbf{p}}{m}+\frac{\hbar}{4m^2c^2}\boldsymbol{\sigma}\times\nabla V\right)u_{\mathbf{k}}=\left(E_{\mathbf{k}}-\frac{\hbar^2k^2}{2m}\right)u_{\mathbf{k}}. \quad (2)$$

The equation for $\mathbf{k}+\mathbf{K}$ is

$$\left[\frac{\mathbf{p}^2}{2m}+V+\frac{\hbar}{4m^2c^2}(\nabla V\times\mathbf{p})\cdot\boldsymbol{\sigma}\right]u_{\mathbf{k}+\mathbf{K}}$$

$$+\hbar\mathbf{k}\cdot\left(\frac{\mathbf{p}}{m}+\frac{\hbar}{4m^2c^2}\boldsymbol{\sigma}\times\nabla V\right)u_{\mathbf{k}+\mathbf{K}}$$

$$+\hbar\mathbf{K}\cdot\left(\frac{\mathbf{p}}{m}+\frac{\hbar}{4m^2c^2}\boldsymbol{\sigma}\times\nabla V\right)u_{\mathbf{k}+\mathbf{K}}$$

$$=\left(E_{\mathbf{k}+\mathbf{K}}-\frac{\hbar^2(\mathbf{k}+\mathbf{K})^2}{2m}\right)u_{\mathbf{k}+\mathbf{K}}. \quad (3)$$

Treating the term

$$\mathcal{H}'=\hbar\mathbf{K}\cdot\left(\frac{\mathbf{p}}{m}+\frac{\hbar}{4m^2c^2}\boldsymbol{\sigma}\times\nabla V\right)$$

as a perturbation, the energy at $\mathbf{k}+\mathbf{K}$ for a nondegenerate level is

$$E_{\mathbf{k}+\mathbf{K}}=E_{\mathbf{k}}+\frac{\hbar^2k^2}{2m}+(\Psi_{\mathbf{k}}|\,\mathcal{H}'|\,\Psi_{\mathbf{k}})+\cdots. \quad (4)$$

Degenerate levels are treated by solving the customary secular determinants.

In working out the matrix elements for perturbation theory, it is helpful to use group theoretical selection rules. Due to the scalar character of the Hamiltonian,

TABLE II. Character table of the double group of Δ; $\mathbf{k}=k[100]$.

Δ	E	$\bar{E}$	$2C_4^2$	$2I\times C_2$	$2I\times C_2'$
$\Delta_1(x)$	1	1	1	1	1
Δ_2	1	1	1	−1	−1
$\Delta_3(y+z)$	1	1	−1	−1	1
$\Delta_4(y-z)$	1	1	−1	1	−1
Δ_5	2	−2	0	0	0

$\Delta_i\times D_{\frac{1}{2}}=\Delta_5$, $i=1,2,3,4$
Δ_5 and Δ_5 are degenerate by time reversal

Selection rules

Δ_i	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5
$\Delta_1\times\Delta_1$	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5
$\Delta_2\times\Delta_2$	Δ_2	Δ_1	Δ_4	Δ_3	Δ_5
$\Delta_3\times\Delta_4$	Δ_4	Δ_3	Δ_3	Δ_1	Δ_5

TABLE III. Character table of the double group of Λ or L; $\mathbf{k}=(k/\sqrt{3})[111]$ or $(\pi/a)(111)$.

Λ, L	E	$\bar{E}$	$2C_3$	$2\bar{C}_3$	$3I\times C_2$	$3I\times\bar{C}_2$
$\Lambda_1(x+y+z)$	1	1	1	1	1	1
Λ_2	1	1	1	1	−1	−1
$\Lambda_3\begin{matrix}(x+\omega y+\omega^2 z)\\(x+\omega^2 y+\omega z)\\(\omega^3=1)\end{matrix}$	2	2	−1	−1	0	0
Λ_4	1	−1	−1	1	i	$-i$
Λ_5	1	−1	−1	1	$-i$	i
Λ_6	2	−2	1	−1	0	0

L_4 and L_5 are degenerate by time reversal
(Λ_4 and Λ_5 are nondegenerate)

Λ_i	Λ_1	Λ_2	Λ_3
$\Lambda_i\times D_{\frac{1}{2}}$	Λ_6	Λ_6	$\Lambda_4+\Lambda_5+\Lambda_6$

Selection rules

Λ_i	Λ_1	Λ_2	Λ_3	Λ_4	Λ_5	Λ_6
$\Lambda_1\times\Lambda_1$	Λ_1	Λ_2	Λ_3	Λ_4	Λ_5	Λ_6
$\Lambda_2\times\Lambda_2$	Λ_2	Λ_1	Λ_3	Λ_5	Λ_4	Λ_6
$\Lambda_3\times\Lambda_3$	Λ_3	Λ_3	$\Lambda_1+\Lambda_2+\Lambda_3$	Λ_6	Λ_6	$\Lambda_4+\Lambda_5+\Lambda_6$

TABLE IV. Character table of the double group of Σ or K; $\mathbf{k}=(k/\sqrt{2})[110]$ or $(3\pi/2a)(110)$.

Σ,K	E	$\bar{E}$	$I\times C_2$	$I\times\bar{C}_2$
$\Sigma_1(z)(x+y)$	1	1	1	1
$\Sigma_2(x-y)$	1	1	−1	−1
Σ_3	1	−1	i	$-i$
Σ_4	1	−1	$-i$	i

$\Sigma_i\times D_{\frac{1}{2}}=\Sigma_3+\Sigma_4, \quad i=1,2$

Selection rules

Σ_i	Σ_1	Σ_2	Σ_3	Σ_4
$\Sigma_1\times\Sigma_i$	Σ_1	Σ_2	Σ_3	Σ_4
$\Sigma_2\times\Sigma_i$	Σ_2	Σ_1	Σ_4	Σ_3

TABLE V. Character table of the double group of X; $\mathbf{k}=(2\pi/a)(100)$.

X	E	$\bar{E}$	$4C_2^2$	$2C_2^{(1)}$	$2I\times C_{2(1)}$	$2I\times\bar{C}_{2(1)}$	$4I\times C_2$
X_1	1	1	1	1	1	1	1
X_2	1	1	1	1	−1	−1	−1
$X_3(x)$	1	1	−1	1	−1	−1	1
X_4	1	1	−1	1	1	1	−1
$X_5(y,z)$	2	2	0	−2	0	0	0
X_6	2	−2	0	0	$\sqrt{2}$	$-\sqrt{2}$	0
X_7	2	−2	0	0	$-\sqrt{2}$	$\sqrt{2}$	0

X_i	X_1	X_2	X_3	X_4	X_5	X_6	X_7
$X_i\times D_{\frac{1}{2}}$	X_6	X_7	X_7	X_6	X_6+X_7		

Selection rules

X_i	X_1	X_2	X_3	X_4	X_5	X_6	X_7
$X_1\times X_i$	X_1	X_2	X_3	X_4	X_5	X_6	X_7
$X_2\times X_i$	X_2	X_1	X_4	X_3	X_5	X_7	X_6
$X_3\times X_i$	X_3	X_4	X_1	X_2	$X_1+X_2+X_3+X_4$	X_6+X_7	X_6+X_7

TABLE VI. Character table of the double group of W; $\mathbf{k}=(2\pi/a)(0\,\frac{1}{2}\,1)$.

W	E	$\bar{E}$	C_2^2	$\bar{C}_2^2$	$I\times C_2$	$I\times\bar{C}_2$	$I\times C_2'$	$I\times\bar{C}_2'$
W_1	1	1	1	1	1	1	1	1
$W_2(x)$	1	1	1	1	−1	−1	−1	−1
$W_3(y+iz)$	1	1	−1	−1	i	i	$-i$	$-i$
$W_4(y-iz)$	1	1	−1	−1	$-i$	$-i$	i	i
W_5	1	−1	$-i$	i	$i\sqrt{i}$	$-i\sqrt{i}$	$\sqrt{i}$	$-\sqrt{i}$
W_6	1	−1	i	$-i$	$\sqrt{i}$	$-\sqrt{i}$	$i\sqrt{i}$	$-i\sqrt{i}$
W_7	1	−1	$-i$	i	$-i\sqrt{i}$	$i\sqrt{i}$	$-\sqrt{i}$	$\sqrt{i}$
W_8	1	−1	i	$-i$	$-\sqrt{i}$	$\sqrt{i}$	$-i\sqrt{i}$	$i\sqrt{i}$

W_i	W_1	W_2	W_3	W_4
$W_i\times D_{\frac{1}{2}}$	W_5+W_6	W_7+W_8	W_5+W_7	W_6+W_8

Selection rules

W_i	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8
$W_1\times W_i$	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8
$W_2\times W_i$	W_2	W_1	W_4	W_3	W_7	W_8	W_5	W_6
$W_3\times W_i$	W_3	W_4	W_2	W_1	W_5	W_6	W_8	W_7
$W_4\times W_i$	W_4	W_3	W_1	W_2	W_6	W_7	W_8	W_5

the term

$$\mathbf{R}=\frac{\hbar}{m}\left(\mathbf{p}+\frac{\hbar}{4mc^2}\boldsymbol{\sigma}\times\boldsymbol{\nabla}V\right)$$

transforms as a vector. Matrix elements of the type $(\Psi_{\mathbf{k}}^{i}|R_x|\Psi_{\mathbf{k}}^{j})$ will vanish unless the direct product $\Gamma_i\times\Gamma_R\times\Gamma_j$ contains the unit representation, where Γ_i is the irreducible representation of wave vector $\mathbf{k}$ according to which $\Psi_{\mathbf{k}}^{i}$ transforms and Γ_R is the representation according to which a vector transforms. Another way of viewing the selection rules is that the only representations that mix with Γ_i under the perturbation $\mathcal{H}'$ are those contained in the decomposition of the direct product $\Gamma_i\times\Gamma_R$. Tables I to VII also give the decomposition of these direct products for the points of symmetry in the Brillouin zone. For convenience the combinations of vector components which transform as a given irreducible representation are indicated in the character tables.

In order to give a more complete treatment of the point Γ, the bases shown in Table VIII may be selected for the irreducible representations.[8] In this notation the spin-orbit splittings at $\mathbf{k}=0$ are, for $\Gamma_4^1-\Gamma_4^3$,

$$\Delta E_{SO}=\frac{3\hbar}{4m^2c^2}i\left(\delta_1\left|\frac{\partial V}{\partial x}p_y-\frac{\partial V}{\partial y}p_x\right|\delta_2\right); \tag{5}$$

and for $\Gamma_5^1-\Gamma_5^4$,

$$\Delta E_{SO}=\frac{3\hbar}{4m^2c^2}i\left(\epsilon_1\left|\frac{\partial V}{\partial x}p_y-\frac{\partial V}{\partial y}p_x\right|\epsilon_2\right). \tag{6}$$

Selection rules suggest that the representations Γ_4 and Γ_5 have first-order matrix elements with $\mathcal{H}'$ and hence finite slopes. Actually all first-order matrix elements with $\mathbf{p}$ vanish due to time reversal symmetry. For example, $(\delta_1|p_y|\delta_3)=(\delta_3|p_y|\delta_1)$ by a reflection in the (101) plane, but by partial integration $(\delta_1|p_y|\delta_3)=-(\delta_3|p_y|\delta_1)$, as the δ_i's are real if the Hamiltonian has time-inversion symmetry; hence all such matrix elements vanish. This argument only holds if all δ_i's are from the same degenerate Γ_4 level.

When spin-orbit interaction is included in the perturbation (i.e., the $\boldsymbol{\sigma}\times\boldsymbol{\nabla}V$ term is not neglected), the energy to first order in k for a Γ_5 level is given by the

TABLE VII. Compatibility relations.

$\Gamma_1\to\Delta_1$	$\Gamma_1\to\Lambda_1$
$\Gamma_2\to\Delta_2$	$\Gamma_2\to\Lambda_2$
$\Gamma_3\to\Delta_1+\Delta_2$	$\Gamma_3\to\Lambda_3$
$\Gamma_4\to\Delta_1+(\Delta_2+\Delta_4)$	$\Gamma_4\to\Lambda_1+\Lambda_3$
$\Gamma_5\to\Delta_2+(\Delta_2+\Delta_4)$	$\Gamma_5\to\Lambda_2+\Lambda_3$
$\Gamma_6,\Gamma_7,\Gamma_8\to\Delta_5$	$\Gamma_6,\Gamma_7\to\Lambda_4$
	$\Gamma_8\to\Lambda_4+\Lambda_5+\Lambda_6$
$X_1\to\Delta_1$	$(L_4+L_5)\to\Lambda_4+\Lambda_5$
$X_2\to\Delta_2$	
$X_3\to\Delta_1$	
$X_4\to\Delta_2$	
$X_5\to(\Delta_3+\Delta_4)$	
$X_6,X_7\to\Delta_5$	

Time reversal degenerate representations are indicated by parentheses.

[8] See F. von der Lage and H. A. Bethe, Phys. Rev. 71, 612 (1947).

secular determinant

$$\begin{vmatrix} -\lambda & (i/2)Ck_+ & -Ck_z & -(i\sqrt{3}/2)Ck_- \\ -(i/2)Ck_- & -\lambda & -(i\sqrt{3}/2)Ck_+ & Ck_z \\ -Ck_z & (i\sqrt{3}/2)Ck_- & -\lambda & (i/2)Ck_+ \\ (i\sqrt{3}/2)Ck_+ & Ck_z & -(i/2)Ck_- & -\lambda \end{vmatrix} = 0, \tag{7}$$

where $\lambda = E_k - (\hbar^2k^2/2m)$, which has solutions

$$\begin{aligned} \lambda &= \pm C\{k^2 + [3(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)]^{\frac{1}{2}}\}^{\frac{1}{2}}, \\ \lambda &= \pm C\{k^2 - [3(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)]^{\frac{1}{2}}\}^{\frac{1}{2}}, \end{aligned} \tag{8}$$

where

$$\begin{aligned} C &= -\frac{1}{2\sqrt{3}}\frac{\hbar^2}{m^2c^2}\left(\delta_1\left|\frac{\partial V}{\partial y}\right|\delta_3\right), \quad \text{for } \Gamma_4^1, \\ C &= -\frac{1}{2\sqrt{3}}\frac{\hbar^2}{m^2c^2}\left(\epsilon_1\left|\frac{\partial V}{\partial y}\right|\epsilon_3\right), \quad \text{for } \Gamma_5^1, \\ C &= 0, \quad \text{for } \Gamma_3. \end{aligned} \tag{9}$$

The first-order energies are shown in Fig. 2 for the [100], [111], and [110] directions. In polar coordinates, $k_x = k\sin\theta\cos\phi$, $k_y = k\sin\theta\sin\phi$, $k_z = k\cos\theta$, Eqs. (8) become

$$\lambda = \pm Ck[1 \pm (3)^{\frac{1}{2}}\sin\theta(\cos^2\theta + \tfrac{1}{4}\sin^2\theta\sin^2 2\phi)^{\frac{1}{2}}]^{\frac{1}{2}}. \tag{10}$$

Figure 3 shows the four first-order energies given by Eq. (10) for constant k plotted against the angle θ for wave vectors in a (110) plane ($\phi = \pi/4$). It is interesting to note that the two-dimensional Λ_6 level is the intersection of surfaces which arise from two different representations in a [100] direction. Any large separation of the energy levels which transform as Γ_8 at the center of the zone, as one proceeds toward the edge of the zone, then would entail a rather complex band structure with highly deformed energy surfaces and accidental degeneracies.[9]

Second-order perturbation theory for the Γ_8 level is tractable on the approximation that only first-order terms in $\sigma\times\nabla V$ and second-order terms in $\mathbf{p}$ be considered. The fourth-order equation which results from the 4×4 secular determinant is

$$\begin{aligned} &y^4 - 2y^3\left[\left(\frac{L-M}{3}\right)^2 k^4 + \frac{N^2-(L-M)^2}{3} \right. \\ &\left. \times(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2) + C^2k^2\right] \\ &+ 4yC^2N(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2) + \left[\left(\frac{L-M}{3}\right)^2 k^4 \right. \\ &\left. + \frac{N^2-(L-M)^2}{3}(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)\right]^2 \\ &+ C^4[k^4 - 3(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)] \\ &+ 2\left(\frac{L-M}{3}\right)^2 C^2(k_x^6 + k_y^6 + k_z^6) \\ &- \left[3\left(\frac{L-M}{3}\right)^2 + \frac{2}{3}N^2\right]C^2k^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2) \\ &+ 21\left(\frac{L-M}{3}\right)^2 C^2k_x^2k_y^2k_z^2 = 0, \end{aligned} \tag{11}$$

where L, M, and N are real numbers and can be expressed in terms of sums over the squares of the

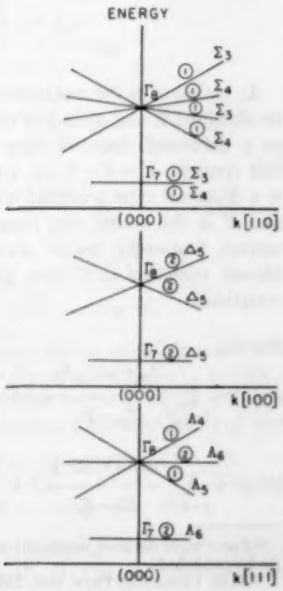


FIG. 2. Plot of energy vs wave vector showing the first-order energy for the spin-orbit split Γ_4 or Γ_5 level in [100], [110], and [111] directions. The circled numbers indicate the dimension of the representation.

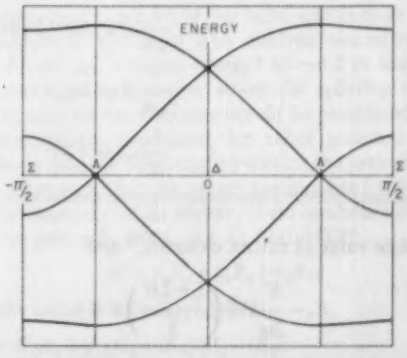


FIG. 3. Cylindrical cross section around the point Γ of a plot of energy vs wave vector for wave vectors in a (110) plane.

[9] C. Herring, Phys. Rev. 52, 365 (1937).

TABLE VIII. Basis functions for the irreducible representations[9] of the double group for the point Γ.

Γ_1, $\alpha=1$	Γ_6, $\alpha\|-\rangle$ $\alpha\|+\rangle$
Γ_2, $\beta=[x^4(y^2-z^2)$ $+y^4(z^2-x^2)$ $+z^4(x^2-y^2)]$	Γ_7, $\beta\|-\rangle$ $\beta\|+\rangle$
Γ_3, $\gamma_1=x^2+\omega y^2+\omega^2 z^2$ $\gamma_2=x^2+\omega^2 y^2+\omega z^2$ $\omega^3=1$	$\Gamma_8(\Gamma_3)$, $\frac{1}{\sqrt{2}}(\omega^2\gamma_2+\omega\gamma_1)\|+\rangle$ $-\frac{i}{\sqrt{2}}(\omega^2\gamma_2-\omega\gamma_1)\|-\rangle$ $\frac{i}{\sqrt{2}}(\omega^2\gamma_2-\omega\gamma_1)\|+\rangle$ $-\frac{1}{\sqrt{2}}(\omega^2\gamma_2+\omega\gamma_1)\|-\rangle$
Γ_4, $\delta_1=x$ $\delta_2=y$ $\delta_3=z$	$\Gamma_7(\Gamma_4^1)$, $\frac{1}{\sqrt{3}}[-i(\delta_1-i\delta_2)\|+\rangle+\delta_3\|-\rangle]$ $\frac{1}{\sqrt{3}}[i(\delta_1+i\delta_2)\|-\rangle-\delta_3\|+\rangle]$
	$\Gamma_8(\Gamma_4^2)$, $\frac{i}{\sqrt{2}}(\delta_1-i\delta_2)\|-\rangle$ $\frac{1}{\sqrt{6}}[i(\delta_1-i\delta_2)\|+\rangle+2\delta_3\|-\rangle]$ $\frac{1}{\sqrt{6}}[i(\delta_1+i\delta_2)\|-\rangle+2\delta_3\|+\rangle]$ $\frac{i}{\sqrt{2}}(\delta_1+i\delta_2)\|+\rangle$
	$\Gamma_7(\Gamma_4^3)$, $\frac{1}{\sqrt{3}}[-i(\epsilon_1-i\epsilon_2)\|+\rangle+\epsilon_3\|-\rangle]$ $\frac{1}{\sqrt{3}}[i(\epsilon_1+i\epsilon_2)\|-\rangle-\epsilon_3\|+>]$
Γ_5, $\epsilon_1=x(y^2-z^2)$ $\epsilon_2=y(z^2-x^2)$ $\epsilon_3=z(x^2-y^2)$	$\Gamma_8(\Gamma_5^3)$, $\frac{i}{\sqrt{2}}(\epsilon_1-i\epsilon_2)\|-\rangle$ $\frac{1}{\sqrt{6}}[i(\epsilon_1-i\epsilon_2)\|+\rangle+2\epsilon_3\|->]$ $\frac{1}{\sqrt{6}}[i(\epsilon_1+i\epsilon_2)\|-\rangle+2\epsilon_3\|+\rangle]$ $\frac{i}{\sqrt{2}}(\epsilon_1+i\epsilon_2)\|+\rangle$

absolute value of matrix elements,[10] and

$$E_k=\frac{\hbar^2}{2m}k^2+\left(\frac{L+2M}{3}\right)k^2+y.$$

For a general $\mathbf{k}$, Eq. (11) has four real roots; for

[10] Dresselhaus, Kip, and Kittel, Phys. Rev. **98**, 368 (1955).

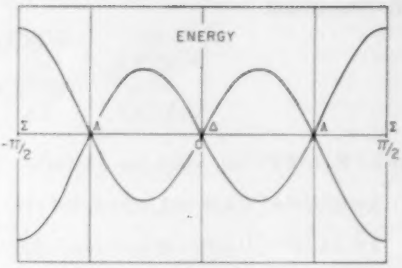


FIG. 4. A cylindrical cross section around the point Γ of a plot of energy vs wave vector for wave vectors in a (110) plane showing the third order splitting of the Γ_6 or Γ_7 surface.

$\mathbf{k}=k[100]$ the energy eigenvalues are

$$E_{k[100]}=\frac{\hbar^2}{2m}k^2+\left(\frac{L+2M}{3}\right)k^2\pm\left[\left(\frac{L-M}{3}\right)^2k^4+C^2k^2\right]^{\frac{1}{2}} \tag{12}$$

where each root is double; and for $\mathbf{k}=(k/\sqrt{3})[111]$

$$E_{(k/\sqrt{3})[111]}=\frac{\hbar^2}{2m}k^2+\left(\frac{L+2M}{3}\right)k^2 + \begin{cases} \frac{|N|}{3}k^2 & \text{(double)} \\ -\frac{|N|}{3}k^2+\sqrt{2}Ck & \\ -\frac{|N|}{3}k^2-\sqrt{2}Ck. & \end{cases} \tag{13}$$

An estimate of the magnitude of the constant C may be obtained if one considers the zinc blende structure as a deformed diamond type structure.[11,12] Here we will consider $V'=V-V_i$ as a perturbation, where V_i is a diamond type potential with inversion symmetry and V is the actual zinc blende type potential. The correct first-order spatial wave functions for a zinc blende structure are then given by the following equations.

For Γ_{1c}

$$\alpha=\alpha^+ + \sum_{i=\Gamma_1^+}\frac{(\alpha_i^+|V'|\alpha^+)}{E_0-E_i}\alpha_i^+ + \sum_{i=\Gamma_2^-}\frac{(\beta_i^-|V'|\alpha^+)}{E_0-E_i}\beta_i^-,$$
$$\alpha=\beta^- + \sum_{i=\Gamma_1^+}\frac{(\alpha_i^+|V'|\beta^-)}{E_0-E_i}\alpha_i^+ + \sum_{i=\Gamma_2^-}\frac{(\beta_i^-|V'|\beta^-)}{E_0-E_i}\beta_i^-. \tag{14}$$

[11] For a more detailed treatment see the paper by F. Herman (to be published).

[12] R. H. Parmenter, Phys. Rev. **100**, 573 (1955).

For Γ_2,

$$\beta=\alpha^-+\sum_{i=\Gamma_1^-}\frac{(\alpha_i^-|V'|\alpha^-)}{E_0-E_i}\alpha_i^-+\sum_{i=\Gamma_2^+}\frac{(\beta_i^+|V'|\alpha^-)}{E_0-E_i}\beta_i^+,$$
$$\beta=\beta^+\sum_{i=\Gamma_1^-}\frac{(\alpha_i^-|V'|\beta^+)}{E_0-E_i}\alpha_i^-+\sum_{i=\Gamma_2^+}\frac{(\beta_i^+|V'|\beta^+)}{E_0-E_i}\beta_i^+. \quad (15)$$

For Γ_3,

$$\gamma_K=\gamma_K^+\sum_{i=\Gamma_{12}^+}\frac{(\gamma_{Ki}^+|V'|\gamma_K^+)}{E_0-E_i}\gamma_{Ki}^+ + \sum_{i=\Gamma_{12}^-}\frac{(\gamma_{Ki}^-|V'|\gamma_K^+)}{E_0-E_i}\gamma_{Ki}^-,$$
$$\gamma_K=\gamma_K^-+\sum_{i=\Gamma_{12}^+}\frac{(\gamma_{Ki}^+|V'|\gamma_K^-)}{E_0-E_i}\gamma_{Ki}^+ + \sum_{i=\Gamma_{12}^-}\frac{(\gamma_{Ki}^-|V'|\gamma_K^-)}{E_0-E_i}\gamma_{Ki}^-. \quad (16)$$

For Γ_4,

$$\delta_K=\delta_K^-+\sum_{i=\Gamma_{15}^-}\frac{(\delta_{Ki}^-|V'|\delta_K^-)}{E_0-E_i}\delta_{Ki}^- + \sum_{i=\Gamma_{25}^+}\frac{(\epsilon_{Ki}^+|V'|\delta_K^-)}{E_0-E_i}\epsilon_{Ki}^+,$$
$$\delta_K=\epsilon_K^+\sum_{i=\Gamma_{15}^-}\frac{(\delta_{Ki}^-|V'|\epsilon_K^+)}{E_0-E_i}\delta_{Ki}^- + \sum_{i=\Gamma_{25}^+}\frac{(\epsilon_{Ki}^+|V'|\epsilon_K^+)}{E_0-E_i}\epsilon_{Ki}^+. \quad (17)$$

For Γ_5,

$$\epsilon_K=\delta_K^+\sum_{i=\Gamma_{15}^+}\frac{(\delta_{Ki}^+|V'|\delta_K^+)}{E_0-E_i}\delta_{Ki}^+ + \sum_{i=\Gamma_{25}^-}\frac{(\epsilon_{Ki}^-|V'|\delta_K^+)}{E_0-E_i}\epsilon_{Ki}^-,$$
$$\epsilon_K=\epsilon_K^-+\sum_{i=\Gamma_{15}^+}\frac{(\delta_{Ki}^+|V'|\epsilon_K^-)}{E_0-E_i}\delta_{Ki}^+ + \sum_{i=\Gamma_{25}^-}\frac{(\epsilon_{Ki}^-|V'|\epsilon_K^-)}{E_0-E_i}\epsilon_{Ki}^-, \quad (18)$$

where the cubic harmonics are as defined in reference 8 except that $\gamma_1^+=\gamma_2^{+*}=x^2+\omega y^2+\omega^2 z^2$, $(\omega^3=1)$, and the $\pm$ superscript is used to denote the parity of the functions.

For the Γ_4 level arising from a diamond type Γ_{25}^+ level

$$C=\frac{1}{\sqrt{3}}\frac{\hbar^2}{m^2c^2}\sum_{i=\Gamma_{15}^-}\frac{(\epsilon_1^+|\partial V_i/\partial y|\delta_{3i}^-)(\delta_{1i}^-|V'|\epsilon_1^+)}{(E_0-E_i)}. \quad (19)$$

To a very rough approximation[10] $V'\sim V_i/Z$ where Z is the atomic number. $C\sim(1/Z)\Delta_{SO}a$, where a is the lattice constant and Δ_{SO} is the spin-orbit splitting at $\mathbf{k}=0$. The first-order energy at the zone edge in a [100] direction is

$$E\sim C\frac{2\pi}{a}\sim 2\pi\frac{\Delta_{SO}}{Z}\sim 0.02 \text{ ev (in InSb)}.$$

Hence the splitting due to the first-order terms would be only of the order of 0.02 ev if the slope were linear all the way to the edge of the zone. Actually the second-order terms which should be large due to the small energy gap will turn the surfaces down very rapidly. Using this value for C and the values of L, M, and N from cyclotron resonance experiments on Ge,[10] it would seem from Eq. (11) that near the center of the Brillouin zone the removal of the twofold Kramers degeneracy is at most 10^{-4} ev for holes with thermal energies. Under these circumstances it seems quite likely that a perturbation expansion about the extremum should contain several orders of perturbation theory, and a simple energy surface with effective mass tensor components independent of wave vector would be a very poor approximation.

For the Γ_6 and Γ_7 level the energy to third order in k is given by

$$E=C_0k^2\pm C_1[k^2(k_x^2k_y^2+k_y^2k_z^2+k_z^2k_x^2)-9k_x^2k_y^2k_z^2]^{\frac{1}{2}}. \quad (20)$$

In third order the levels are split in all but the [100] and [111] directions. In polar coordinates Eq. (20) reads

$$E=C_0k^2\pm C_1k^3\sin\theta[1-\sin^2\theta(1+2\sin^2 2\phi)+(9/4)\sin^2 2\phi\sin^4\theta]^{\frac{1}{2}}. \quad (21)$$

Figure 4 shows a plot of energy *vs* θ for constant k and $\phi=\pi/4$ [i.e., $\mathbf{k}$ in a (110) plane]. The surfaces have their maximum separation along <110> axes. In the III-V class of semiconducting compounds like InSb, the high mobility electrons are presumably in a spherically symmetric Γ_6 state.[4] Higher orders in k should not enter until the thermal or Fermi energy for the electrons is of the order of the band splittings at $k=0$. In impure n-type InSb with electron concentrations $\sim 10^{18}$/cm^3 and a Fermi energy[12] of $\sim$0.2 ev one may be entering into a region where the splitting of the degeneracy for the electrons should be considered.

Perturbation expansions for other points in the Brillouin zone are facilitated by writing the perturbation in a form such that the vector combinations indicated in the character tables appear; these combinations are, for the points Γ, (000), and X, $(2\pi/a)(100)$,

$$3\mathcal{C}'=k_xR_x+k_yR_y+k_zR_z; \quad (22)$$

for the point W, $(2\pi/a)(0,\frac{1}{2},1)$,

$$3\mathcal{C}'=K_xR_x+\tfrac{1}{2}[(K_y+iK_z)(R_y-iR_z)+(K_y-iK_z)(R_y+iR_z)]; \quad (23)$$

[12] Hrostowski, Wheatley, and Flood, Phys. Rev. 95, 1683 (1954).

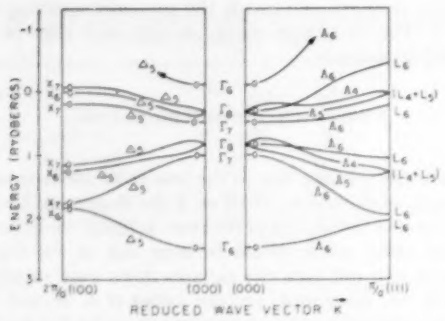


FIG. 5. A schematic drawing of the energy levels for a zinc blende type structure modification of boron nitride based on Herman's calculation for diamond [F. Herman, Phys. Rev. **88**, 1210 (1952); thesis, Columbia University, 1953 (unpublished)]. The spin-orbit splittings are highly exaggerated for the purpose of illustration. The levels marked with an 0 have zero slope along that axis. This figure should be compared with Elliott's Fig. 2. [R. J. Elliott, Phys. Rev. **96**, 266 (1954).]

for the points Λ, $(k/\sqrt{3})(111)$, and L, $(\pi/a)(111)$,

$$\mathcal{H}' = \tfrac{1}{3}[(K_x+K_y+K_z)(R_x+R_y+R_z) + (K_x+\omega K_y+\omega^2 K_z)(R_x+\omega^2 R_y+\omega R_z) + (K_x+\omega^2 K_y+\omega K_z)(R_x+\omega R_y+\omega^2 R_z)], \quad (24)$$

and for the points Σ, $(k/\sqrt{2})(110)$, and Δ, $k(001)$

$$\mathcal{H}' = K_z R_z + \tfrac{1}{2}[(K_x+K_y)(R_x+R_y) + (K_x-K_y)(R_x-R_y)]. \quad (25)$$

Using these perturbations the first-order energies can be written down at sight and are of the following form: for Δ_5, $k(100)$,

$$E = C_2 K_x \pm C_3(K_y^2+K_z^2)^{\frac{1}{2}}; \quad (26)$$

for Λ_4 and Λ_5, $(k/\sqrt{3})(111)$,

$$E = C_4(K_x+K_y+K_z); \quad (27)$$

for Λ_6,

$$E = C_5(K_x+K_y+K_z) \pm C_6[K^2-(K_xK_y+K_yK_z+K_xK_z)]^{\frac{1}{2}}; \quad (28)$$

for Σ_4 or Σ_5, $(k/\sqrt{2})(110)$,

$$E = C_7(K_x+K_y)+C_8K_z; \quad (29)$$

and for X_6 or X_7, $(2\pi/a)(100)$,

$$E = \pm C_9(K_y^2+K_z^2)^{\frac{1}{2}}. \quad (30)$$

The second-order energies for the point W, $(2\pi/a)(0,\frac{1}{2},1)$, are of the form

$$E = C_{10}K_x^2 + C_{11}(K_y^2+K_z^2). \quad (31)$$

If the energy extremum is at the point W, the constant energy surfaces will be spheroids with $<100>$ axes.

The irreducible representations Λ_4 and Λ_5 only have slopes along a $<111>$ axis, hence if the constant C_4 is zero at some point along the axis as it presumably is near the center of the zone in InSb due to the perturbing influence of the lowest conduction states (see Fig. 5), then the constant energy surfaces at these extrema will be spheroids with $<111>$ axes. However, because of the smallness of the first-order terms compared to the second-order terms, it is not expected that these extrema will be of any significance in the band structure if they occur near the center of the zone.

TABLE IX. Possible energy extrema for the representations of the double group in zinc blende structures.

Extrema at		Representations of double group	Constant energy surfaces
(000)	Γ	Γ_8	sphere
		Γ_7	sphere
		Γ_8 (from Γ_8)	2 warped energy surfaces as for holes in Si and Ge
$(k/\sqrt{3})[111]$	Λ	Λ_4	8 spheroids; $<111>$ axes
		Λ_5	8 spheroids; $<111>$ axes
$(2\pi/a)(1,\frac{1}{2},0)$	W	W_5, W_6, W_7, W_8	6 spheroids; $<100>$ axes
General point		...	48 general ellipsoids

The representation Γ_8 at the center of the zone can have zero slope if it arises from the two-dimensional representation Γ_3 of the single group. In this case the energy to second order in k is given by

$$E = C_{12}k^2 \pm [C_{13}^2k^4 + C_{14}^2(k_x^2k_y^2+k_y^2k_z^2+k_x^2k_z^2)]^{\frac{1}{2}}, \quad (32)$$

in the limit that all spin-orbit splittings are negligible compared with the spacing between levels at $k=0$, then $C_{14}^2 = -3C_{13}^2$. In atomic Sn it is known that the $4d$ levels overlap the $5s$ atomic levels, hence in grey Sn or InSb it is conceivable that a Γ_3 level, which can be represented in a tight binding approximation by d-orbitals, could be the uppermost valence band.

Table IX gives a tabulation of the types of energy surfaces which might be expected at certain points in the Brillouin zone. It should be emphasized in all these considerations that if the spin-orbit interaction is small or the difference in the crystal potential from the diamondlike potential is only slight, then the region of convergence of the energy expressions will be small compared to kT and one should then consider only the single group representations in the first case, or the diamond structure double group representations in the latter case. If the region of convergence of the perturbation expansion is $\sim kT$, then one is not justified in keeping only the lowest nonvanishing term and a simple effective mass approximation would seem unjustified. From the order of magnitude estimate of the spin-orbit splitting, it seems this might be the case for holes in InSb.

I should like to thank Professor C. Kittel for suggesting this problem and for guidance throughout the course of the investigation. I am indebted to Dr. F. Herman and Dr. R. Paramenter of R.C.A. Laboratories for communicating their results to me prior to publication. I also wish to thank Mr. R. E. Behringer for checking the manuscript.

Fluorescence Excitation Spectrum of Anthracene

G. T. WRIGHT*
Physics Department, Rhodes University, Grahamstown, South Africa
(Received January 28, 1955)

Surface escape of fluorescence is enhanced in crystals which strongly re-absorb their own fluorescence. This is investigated quantitatively for anthracene. The apparent fluorescence efficiency, measured by viewing the intensity of fluorescence transmitted through the crystal, is found to decrease as the depth of penetration of the excitant decreases. For photo-excitation this produces a fluorescence excitation spectrum since different exciting wavelengths penetrate different distances. The minima of the excitation spectrum correspond to absorption maxima. The maxima of the first absorption band of the crystal are hence found to occur at wavelengths 3960 A, 3740 A, 3540 A, 3370 A, and 3220 A.

INTRODUCTION

MANY organic crystals are opaque to part of their own emission and consequently the characteristics of the molecular fluorescence are considerably modified in the fluorescence from the crystal. This has been investigated quantitatively for several organic crystals[1–3] and is particularly appreciable for anthracene crystals which re-absorb about three-quarters of the molecular fluorescence.

A consequence of self-absorption is that, under conditions of surface excitation, it increases the fraction of fluorescence which escapes through the crystal surface.[4] Thus the apparent fluorescence efficiency of the crystal, measured by viewing the intensity of fluorescence transmitted through the crystal, is decreased. In addition, surface escape of fluorescence is influenced by the depth of penetration of the excitant being smallest for a deeply penetrating excitant. This is because a considerable part of the fluorescence emitted towards the crystal surface from excited molecules produced initially deep within the crystal is re-absorbed before it can reach the surface and escape. Experiments have been made to investigate this aspect of surface escape of fluorescence and preliminary results for anthracene crystals are reported here.

EXPERIMENTAL

Laminae, approximately 10×10×2 mm, were cleaved from scintillation grade crystals obtained from the Larco Nuclear Instrument Company and mounted above an RCA type 5819 photomultiplier. A diaphragm was placed beneath the crystal to ensure that the photomultiplier received only fluorescence transmitted through the crystal. The upper surface of the crystal was illuminated with wavelengths in its absorption region using monochromatic light from a Beckman Model D.U. Spectrophotometer. The relative quantum intensity of the exciting light was measured at each wavelength with a Hilger type F.T. 16 linear vacuum thermopile fitted with a fluorite window. The measured intensity of the transmitted fluorescence was hence normalized to constant quantum intensity of the exciting light. By plotting the normalized fluorescence intensity as a function of exciting wavelength the fluorescence excitation spectrum of anthracene was obtained. This is shown in Fig. 1 for excitation normal to the (001) crystal plane.

DISCUSSION

The similarity of the excitation spectrum to the inverted absorption spectrum is evident and is expected since the intensity variations are due to the variations in extinction coefficient of the crystal. To increase surface escape effects, the freshly cleaved crystal surface

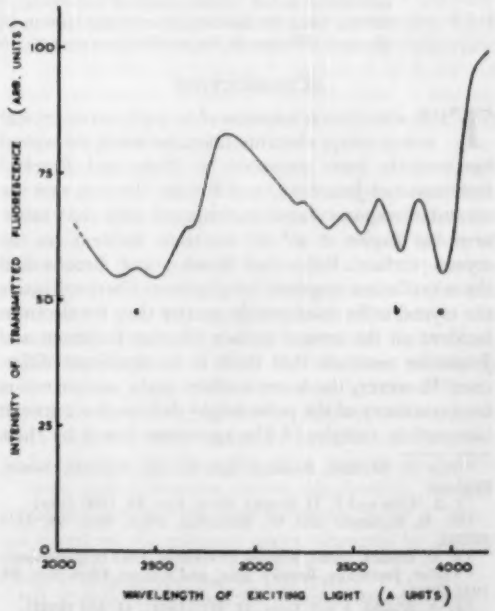


FIG. 1.

* Now at Mullard Research Laboratories, Salfords, Surrey, England.

[1] J. B. Birks and W. A. Little, Proc. Phys. Soc. (London) **A66**, 921 (1953).

[2] G. T. Wright, South African Congress of Physics, Pretoria, July, 1954 (unpublished).

[3] J. B. Birks and G. T. Wright, Proc. Phys. Soc. (London) **B67**, 657 (1954).

[4] G. T. Wright, Proc. Phys. Soc. (London) **B68**, 241 (1955).

was scraped with a clean blade to produce a translucent appearance. A direct measurement showed that variations in the reflection coefficient of the surface were very small (≃5%) and acted to oppose the variations in the excitation spectrum. No correction was made for them.

Since the minima of the fluorescence excitation spectrum correspond to absorption maxima this curve provides much useful information about the absorption spectrum of the crystal. This is difficult to measure directly because of the high extinction coefficients involved. For the first absorption band the absorption maxima are observed to occur at wavelengths 3960 A, 3740 A, 3540 A, 3370 A, and 3220 A; the mean wave number spacing of the vibrational structure is 1430 cm^{-1}. These positions of the absorption maxima are at slightly longer wavelengths than previously reported by Kortum and Finckh.[5] The present measurements were made with large single crystals, however, and are therefore more indicative of the true absorption spectrum of the crystal.

[5] G. Kortum and B. Finckh, Z. physik. Chem. **B52**, 263 (1942).

CONCLUSIONS

For substances such as anthracene, which possess sufficient overlap of fluorescence and absorption spectra, the fluorescence excitation spectrum offers a means of investigating the absorption spectrum of the crystal. The measurements may be extended into the vacuum ultraviolet and soft x-ray regions and may provide valuable information about higher electronic states and ionization potentials of molecules in the solid state.

ACKNOWLEDGMENTS

The author wishes to thank the South African Council for Scientific and Industrial Research for a research grant to purchase equipment.

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Scintillation Response of Anthracene Crystals to Short Range Electrons

G. T. WRIGHT*
Physics Department, Rhodes University, Grahamstown, South Africa
(Received January 28, 1955; revised manuscript received June 6, 1955)

The scintillation response of anthracene crystals to short range electrons is discussed. There is a significant difference between the response to electrons incident on the crystal surface and to electrons liberated inside the crystal. Possible explanations for this difference are discussed. The "primary" photon cascade theory proposed by Birks is considered unacceptable particularly in view of recent measurements of the shapes of scintillation pulses. Surface escape of fluorescence, proposed previously by the author, is discussed quantitatively using the fluorescence excitation spectrum of anthracene and is considered adequate to explain the observed difference in the scintillation response curves.

INTRODUCTION

THE scintillation response of an anthracene crystal to low-energy electrons liberated inside the crystal has recently been measured by Birks and Brooks,[1] Robinson and Jentschke,[2] and Brooks.[3] In each case the measured response curve is compared with that measured by Taylor *et al.*[4] for electrons incident on the crystal surface. Birks and Brooks, and Brooks find the scintillation response for electrons liberated inside the crystal to be considerably greater than for electrons incident on the crystal surface whereas Robinson and Jentschke conclude that there is no significant difference. However, the latter authors make no correction for asymmetry of the pulse-height distribution curve at low-particle energies.[5,6] The agreement found by them between the two response curves is thus fortuitous since this correction raises their internal response curve above that for surface incident electrons. It is apparent, then, that although the response curves measured by these various authors for internally liberated electrons do not agree very well the scintillation efficiency of anthracene is greater for these electrons than for surface incident electrons. This difference is discussed in the present communication.

DISCUSSION

An explanation for this difference has been given by Birks[7] who considers that for external electrons surface escape of postulated "primary" photons reduces the scintillation efficiency below that for internal electrons. These primary photons are due to allowed transitions

* Now at Mullard Research Laboratories, Salfords, Surrey, England.

[1] J. B. Birks and F. D. Brooks, Phys. Rev. **94**, 1800 (1954).
[2] W. H. Robinson and W. Jentschke, Phys. Rev. **95**, 1412 (1954).
[3] F. D. Brooks, thesis, Rhodes University, 1955 (unpublished).
[4] Taylor, Jentschke, Remley, Eby, and Kruger, Phys. Rev. **84**, 1034 (1951).
[5] G. T. Wright, J. Sci. Instr. **31**, 377 (1954); **31**, 462 (1954).
[6] G. T. Wright, Phys. Rev. **96**, 569 (1954). The scintillation response of anthracene and *trans*-stilbene crystals to low-energy photoelectrons has recently been measured by J. M. Fowler and C. E. Roose [Phys. Rev. **98**, 996 (1955)] who find the scintillation intensity to be proportional to the electron energy over the measured range from 10 kev to 40 kev. This agrees with the response curve found earlier by the author[6] on correcting the results of Birks and Brooks' for the effects of photomultiplier statistics.
[7] J. B. Birks, *Scintillation Counters* (Pergammon Press, London; McGraw-Hill Book Company, Inc., New York, 1953); Phys. Rev. **94**, 1567 (1954).

from a highly excited electronic state of the molecule to the ground state and possess energies of the order of the molecular ionization energy (≃8 ev). The molecules which give rise to these photons are those which are initially ionized by the passage of the incident electron. Those primary photons which do not escape through the crystal surface undergo a rapid cascade process of reabsorption and re-emission in the crystal during which their energy is degraded to the level of the first electronic excited state and from which normal fluorescence occurs. Although this process, with certain other assumptions, can qualitatively explain a wide range of fluorescence and scintillation phenomena, no direct evidence has been obtained for the existence of primary photons or of allowed electronic transitions from excited levels higher than the first. Direct observation of the postulated primary photons is difficult since they lie in the vacuum ultraviolet, but if radiative transitions from high excited states are allowed it should be possible to observe fluorescence emission at all wavelengths up to that of the exciting light by observing reflected fluorescence from microcrystals or from the surface of a single crystal. Excitation of anthracene with electrons[8] and with 2537 A radiation,[9] however, yields only the normal fluorescence spectrum corresponding to transitions from the first electronic excited level to the ground state. If the primary photons exist it is possible that the energy degradation through the main absorption bands of the crystal occurs by a nonradiative process but to this author it does not seem justifiable to assume the existence of these high-energy photons (which should not be confused with soft x-ray emission) without some direct experimental evidence. Further, this description of the scintillation process does not take into account fluorescence emission from molecules which are raised only into low electronic levels by the passing particle. The number of such molecules will certainly be greater than the number ionized and cannot be neglected. In addition, explanation in terms of a photon cascade process predicts an exponential decay for the scintillation pulse. Recent measurements in the author's laboratory (to be published) show that for both electron and alpha-particle excitation the scintillation consists of a short initial spike of fluorescence followed by a longer decay component; the majority of the emission is contained in the long decay component.[10] This observation supports the description previously given by the author of quenching processes in the primary excitation column.[11]

An alternative explanation for the different scintillation response of anthracene to internal and external electrons has been given by the author who has suggested that this is due to surface escape of fluorescence.[6] This possibility has been discussed by Birks but is discarded by him.[7] However, recent experimental data for anthracene supports this explanation.

An investigation of the effects of self-absorption on fluorescence in organic crystals[12] has shown that under conditions of surface excitation surface escape of fluorescence is enhanced if a crystal reabsorbs an appreciable part of its own fluorescence. As a consequence the apparent fluorescence efficiency of the crystal (measured by viewing the fluorescence intensity transmitted through the crystal) decreases as the depth of penetration of the excitant decreases.[13] For surface incident electrons this produces a decrease in the scintillation efficiency over and above any decrease due to quenching mechanisms in the primary excitation column. This is the cause of the difference between the scintillation response of anthracene to internal and external electrons; its effects may be estimated quantitatively for anthracene and compared with the available experimental data.

Using the absorption spectrum of crystalline anthracene measured by Kortum and Finckh,[14] the mean penetration depths of various wavelengths may be evaluated. Using the fluorescence excitation spectrum[13] the apparent reduction in fluorescence efficiency, due to surface escape, may be related to the mean penetration depth of the exciting light. In this way a quantitative relation is obtained between the penetration depth of the excitant and the apparent fluorescence efficiency. To use this for electrons it is necessary to know the mean penetration depths of the electrons. These are difficult to estimate because of the multiple scattering suffered by low-energy electrons. Approximate calculations have been made by Brooks based on the theory of scattering in thin foils and have been used here. These calculations are sufficiently accurate for the present purpose since the extent of surface escape of fluorescence is very insensitive to the absolute penetration depth of the excitant; the fluorescence excitation spectrum shows that a hundredfold reduction in the penetration depth reduces the apparent fluorescence efficiency less than twofold. For the same reason we may neglect slight differences due to the different excitation densities produced by photons and electrons and compare directly surface escape effects for photons and electrons with the same mean penetration depths.

Some typical response curves calculated in this way for surface incident electrons are shown in Fig. 1. These are based on the response curve measured by Brooks

[8] W. A. Little, thesis, Rhodes University, 1955 (unpublished).
[9] J. B. Birks and G. T. Wright, Proc. Phys. Soc. (London) **B67**, 657 (1954).
[10] These decay time measurements were made for scintillations from crystals of anthracene, *trans*-stilbene and *para*-terphenyl by measuring the power developed by the scintillation current in the output load of the multiplier. For a steady current the power is i^2R but when the current is delivered in short pulses the power level increases. For the particular case of smooth exponential pulses with decay time τ and repetition frequency z the power developed is $i^2R/2z\tau$. Hence by measuring the power level the scintillation decay time may be obtained. A detailed account of these measurements is being published elsewhere.
[11] G. T. Wright, Phys. Rev. **91**, 1282 (1953).
[12] G. T. Wright, Proc. Phys. Soc. (London) **B68**, 241 (1955).
[13] G. T. Wright, preceding paper [Phys. Rev. **100**, 587 (1955)].
[14] G. Kortum and B. Finckh, Z. Phys. Chem. **B52**, 263 (1942).

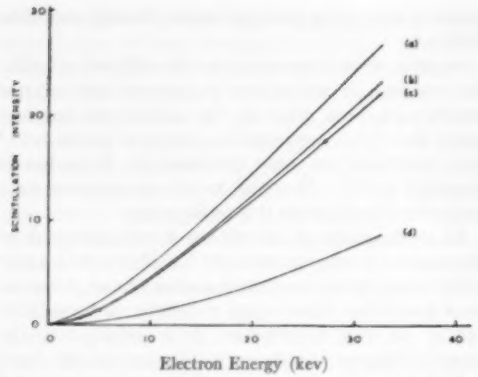


FIG. 1. Scintillation response of anthracene to surface-incident electrons. Curves (a), (b), (d): computed from surface-escape effects; (a), (b): freshly cleaved surfaces; (d): polished surface of 1 cm² scintillation grade crystal. Curve (c): experimental response (Taylor *et al.*).

for internal electrons. Curves (a) and (b) refer to freshly cleaved surfaces; the former was given by a smooth, clear surface and the latter by a translucent surface. Curve (d) was given by a smooth polished surface aged in the dark by exposure to the atmosphere for several months. The very low efficiency of this surface is attributed to chemical deterioration during ageing and is being investigated further. The experimental response curve for surface incident electrons measured by Taylor *et al.* is shown by curve (c); a small correction, about 5%, has been applied to correct for back diffusion of incident electrons.[16] The state of the crystal surface during this measurement is not known but since the crystal was mounted in vacuum it is probable that sublimation produced a slightly translucent surface which would enhance surface escape effects.

[16] E. J. Sternglass, Phys. Rev. **95**, 345 (1954).

CONCLUSIONS

The scintillation response of anthracene to surface incident electrons is very sensitive to surface effects particularly if the crystal surface is not smooth and freshly cleaved. These effects are due to surface escape of fluorescence and to chemical deterioration of the crystal surface and are sufficient to explain the observed difference in the response of anthracene to internal and external electrons. It is desirable, however, to measure the scintillation response and the fluorescence excitation spectrum for the same crystal surface in order to ascertain the precise part played by surface effects.

Photofluorescence measurements made with *para*-terphenyl show surface effects to be absent. This is as expected since re-absorption in this crystal is negligible. Accordingly, for this material the scintillation response should be the same for internal and external electrons apart from slight differences due to back diffusion of surface incident electrons.

For the heavy particles surface effects must be considered at low energies. For alpha particles absorbed in anthracene they become appreciable at energies less than a few Mev and provide a possible explanation for the saturation of the quantity dL/dX observed at high values of dE/dX.

ACKNOWLEDGMENTS

The author wishes to thank Mr. F. D. Brooks for communicating unpublished data and the South African Council for Scientific and Industrial Research for a research grant to purchase equipment.

Surface Potential and Surface Charge Distribution from Semiconductor Field Effect Measurements

W. L. BROWN
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received April 28, 1955)

It is possible from measurements of the change in conductance of a semiconductor with application of an electric field normal to its surface to determine both the electrostatic potential of the surface and the distribution of charge in surface states. Such determinations depend upon the uniqueness of the minimum in conductance which can be observed in these experiments, and its independence of the surface state charge.

MEASUREMENTS of the change in conductance of semiconductor samples with application of a transverse electric field have previously indicated[1] a dependence on the properties of the semiconductor surface. It has recently become apparent that one can obtain from such experiments a direct determination of the surface potential (the electrostatic potential of the surface relative to the interior of the semiconductor) and quantitative information concerning the distribution of charge in nonconducting electron states in the energy gap at the semiconductor surface. The essential feature of these determinations is the recognition that as the surface potential is varied the conductance of the sample will go through a minimum which has a unique value independent of any immobile surface charge. When the minimum is observed experimentally, it serves as a reference by which the measured conductance as a function of the applied electric field can be related to the surface potential and the charge in surface states.

[1] W. Shockley and J. L. Pearson, Phys. Rev. **74**, 232 (1948).

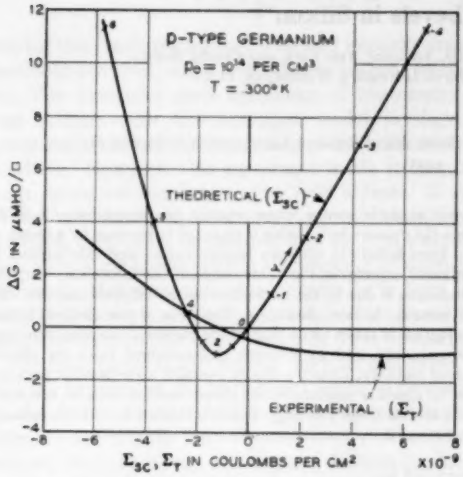


FIG. 1. Theoretical dependence of surface conductance on space charge and experimentally determined dependence on total charge.

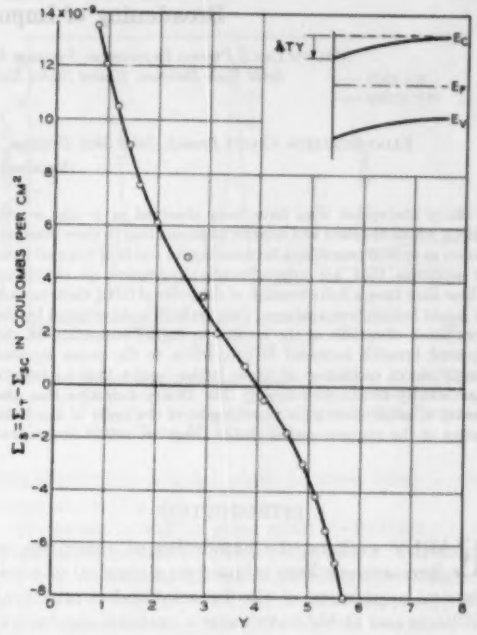


FIG. 2. Charge in surface states as a function of surface potential.

The total charge on the semiconductor per unit of surface area is $\Sigma_T = kE/4\pi$, where E is the applied electric field. This is divided as $\Sigma_T = \Sigma_{sc} + \Sigma_s$, where Σ_{sc} is the charge in the space-charge region adjacent to the surface but inside the semiconductor and Σ_s is the charge in surface states, assumed to be immobile. The conductance of the sample depends only upon Σ_{sc}, and one can calculate this dependence as Garrett and Brattain[2] have done. Such a theoretical curve is shown in Fig. 1 for a high-resistivity p-type germanium sample at 300°K. ΔG is the change in conductance relative to the conductance at $\Sigma_{sc}=0$. For positive Σ_{sc}, the hole concentration near the surface is increased with a corresponding increase in the sample conductance. For Σ_{sc} small and negative, both the hole concentration and the conductance decrease. For Σ_{sc} large and negative, an inversion layer of high electron concentration is formed at the surface, increasing the conductance again and producing a minimum in the curve. The slopes on the theoretical curve approach μ_p for Σ_{sc} large and positive and μ_n for Σ_{sc} large and negative. (No correction has been applied for reduction in mobility due to constraint of the carriers to a thin surface region. According to Schrieffer's analysis,[3] in the particular case considered here the correction would be small.) Indicated along the theoretical curve are values of V, the surface potential expressed in units of kT/q.

Figure 1 also shows an experimental curve taken with a 180-cps external electric field on an etched sample in dry nitrogen. The abscissa is now Σ_T. The experimental curve has been adjusted *vertically* so that its minimum occurs at the minimum of the theoretical curve. This adjustment is significant since the minimum conductance has a unique value independent of Σ_s. Having made this adjustment, the experimental conductance has an absolute significance, in terms of Σ_{sc} and V. The difference between the experimental and theoretical curves at a given value of ΔG is just Σ_s. Thus by comparing the two curves one can obtain Σ_s as a function of V. Figure 2 is a plot of this result. As indicated, the surface states are neutral when $V \simeq 4$, i.e., when the Fermi level is about $3kT$ above the middle of the band gap at the surface. (For this sample the Fermi level in the body is about $1kT$ below the middle of the band.) The shape of the curve is related to, but does not uniquely determine, the energy distribution of the surface states. The distribution may be continuous, in which case the slope of the curve is the surface state density involving the minimum number of states; or the distribution may be discrete, perhaps as described by Bardeen and Brattain[4] in their model of the surface. This later possibility would predict a curve similar in shape to Fig. 2.

Referring again to Fig. 1 at $\Sigma_T=0$, the measured conductance indicates a value of $V=+3$. The chemical state of this surface is such that in the absence of an external electric field it has a surface potential of +0.08 volt.

In order to make the foregoing determinations it is essential to obtain the conductance minimum in the experimental curve. This can always be done, at least in principle, with large enough electric fields. Within this limiation, it is thus possible to study the surface potential and the distribution of charge in surface states by field-effect measurements alone.

[2] C. G. B. Garrett and W. H. Brattain, Phys. Rev. **99**, 376 (1955).
[3] J. R. Schrieffer, Phys. Rev. **94**, 1420 (1954).
[4] W. H. Brattain and J. Bardeen, Bell System Tech. J. **32**, 1 (1953).

Broadening of Impurity Levels in Silicon*†

MELVIN LAX,‡ *Physics Department, Syracuse University, Syracuse, New York, and Crystal Branch, Solid State Division, United States Naval Research Laboratory, Washington, D. C.*

AND

ELIAS BURSTEIN, *Crystal Branch, Solid State Division, United States Naval Research Laboratory, Washington, D. C.*

(Received July 12, 1955)

Sharp absorption lines have been observed in p- and n-type silicon whose absolute and relative positions lead to their interpretation as optical transitions between bound states of trapped holes or electrons that are approximately hydrogenic in character. These lines have a finite breadth of the order of 0.001 electron-volt at liquid helium temperatures. This breadth is determined by the zero-point vibrations of the lattice. At higher temperatures, the squared breadth increases in proportion to the mean squared amplitude of oscillation of those lattice modes that contribute significantly to the broadening. The theory indicates that the modes of importance have wavelengths of the order of the Bohr radius of the trapped carrier state. These are rather long wavelength acoustic modes whose energy $\hbar\omega$ corresponds to 80°K. Thus the squared broadening is expected to increase by a factor of two from helium to nitrogen temperatures—and this increase is confirmed experimentally. This confirms the hypothesis that the broadening is due to the interaction of the trapped electron with the acoustic lattice vibrations. The form of the electron-lattice interaction is taken to be that of the Bardeen-Shockley deformation potential, and its strength is determined from the experimental mobility. Thus the theory contains no adjustable parameters. In absolute magnitude, the theoretical line breadth turns out to be several times too large. Possible reasons for the discrepancy are discussed.

1. INTRODUCTION

EARLY work on the broadening of absorption or luminescence lines is based on a classical or semiclassical application of the Franck-Condon principle.[1] Williams and Hebb[2] have made a quantum mechanical treatment of luminescence in thallium activated potassium chloride using a single configurational coordinate. Huang and Rhys[3] in treating F-center absorption made the first quantum mechanical treatment of a many-coordinate problem. Their technique is limited, however, by the requirement that all the normal coordinates have the same frequency.

The present problem—broadening of impurity levels in homopolar semiconductors—cannot be treated by any of the previously available methods: (1) The criterion previously developed for the validity of the semiclassical Franck-Condon principle[1] will be shown to be violated [see Eq. (5.37)]. (2) The orbits of impurity trapped electrons are so large that interaction takes place with many ($\sim 10^2$) atoms[4] and a Williams-type treatment based on a single real coordinate is inadequate. (3) Acoustical modes will be more important for broadening at all temperatures below the Debye temperature. The spread in frequencies in these modes prohibits the use of the Huang-Rhys analysis.

A generalization of the Huang-Rhys method to the case of modes with an arbitrary frequency distribution was therefore developed.[5] The relationship between this generalized procedure, the Huang-Rhys method and the more recent modification introduced by O'Rourke[6] will be discussed in Sec. 2. In this section, we shall attempt to summarize the physical assumptions made in carrying out the calculation.

The chief approximation made in the course of our calculation has been that we have dealt with the broadening of impurity levels of a hydrogenic type. A hydrogenic energy level scheme will not be quite right unless one has a single nondegenerate band whose band edge is at $k=0$, so that a simple effective mass theory is valid. It is now clear from cyclotron resonance experiments[6] that the valence band of both silicon and germanium is degenerate, and the conduction bands have minima at points other than $k=0$. At the time the calculations reported here were made, however, this information was not available, and the simplest possible assumptions were made. It seems worth while to report these calculations in their original form because they illustrate the essential mechanism for broadening without leading to undue complications of a purely computational nature.

The results will be compared with experiments on boron-doped silicon (see Figs. 1 and 2) because the most reliable and complete experimental information is avail-

* A summary of the results contained here was reported to the American Physical Society in March, 1953 [Phys. Rev. **91**, 208 (1953)].

† Supported by the Department of the Navy (Office of Naval Research) and the Department of the Air Force (Office of Scientific Research). Publication of this document, for any purpose, by the United States Government, is authorized.

‡ Present address: Bell Telephone Laboratories, Murray Hill, New Jersey, on leave of absence from Syracuse University.

[1] For references to earlier work and a critical analysis of the Franck-Condon principle see M. Lax, J. Chem. Phys. **20**, 1752 (1952) hereafter referred to as I. See also reference 14.

[2] F. E. Williams and M. H. Hebb, Phys. Rev. **84**, 1181 (1951).

[3] K. Huang and A. Rhys, Proc. Roy. Soc. (London) **A204**, 406 (1950).

[4] The number of atoms in a Bohr sphere is $(4\pi a^3/3)(2n_0)\simeq 500$ atoms where $n_0=2.5\times 10^{22}$ primitive cells per cm³, 2=number of atoms per primitive cell, and a, the Bohr radius of the impurity state is shown in Eq. (1.2) to be about 13.5 A for p-type silicon doped with boron.

[5] R. C. O'Rourke, Phys. Rev. **91**, 265 (1953).

[6] Dresselhaus, Kip, and Kittel **92**, 827 (1953); Lax, Zeiger, Dexter, and Rosenblum, Phys. Rev. **93**, 1418 (1954); R. N. Dexter and B. Lax, Phys. Rev. **96**, 223 (1954); Dexter, Lax, Kip, and Dresselhaus, Phys. Rev. **96**, 222 (1954).

able for this case[7]: transitions to excited impurity states have been observed, and a level scheme has been worked out. The unusually good agreement of the empirical level scheme with the hydrogen model is slightly fortuitous in view of the degeneracy of the valence band in silicon. However, this agreement made it easy to assign quantum numbers to the level scheme. If the correct wave functions for acceptor impurity states were available, they could be inserted into (3.1) to yield an estimate for the broadening in the presence of degenerate bands. However, only donor states have been investigated theoretically in any detail, and for these the energy levels are better known than the wave functions.[8]

The basic physical assumption made in this paper is that the broadening of the impurity levels is produced by the lattice vibrations through the electron-lattice interaction. Using the Born-Oppenheimer viewpoint, we may say that the energy of a trapped electron (or hole) is a function of the nuclear coordinates. The energy difference between a ground and excited impurity state will depend on the nuclear configuration. The frequency of the absorbed radiation will then vary as the nuclei oscillate, thus leading to a broadening of the absorption line. The mean square width of the line will then be proportional to the mean square amplitude of nuclear oscillations. Thus, at absolute zero there will be a zero-point width associated with the zero-point oscillations of the lattice. The squared width will increase gradually with temperature. At temperatures at which the classical equipartition law becomes valid, $kT \gtrsim \hbar\omega$ (where ω is a frequency of a typical mode of importance in the broadening) the squared width becomes linear in the temperature (Fig. 3).

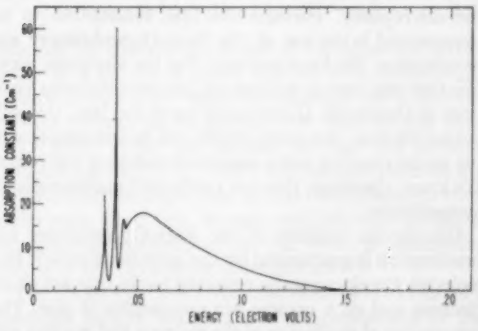


FIG. 1. Infrared absorption of a single crystal of silicon doped with boron [Burstein, Bell, and Henvis (unpublished)]. Further discussion of this data may be found in Burstein, Bell, Davisson, and Lax, J. Phys. Chem. **57**, 849 (1953).

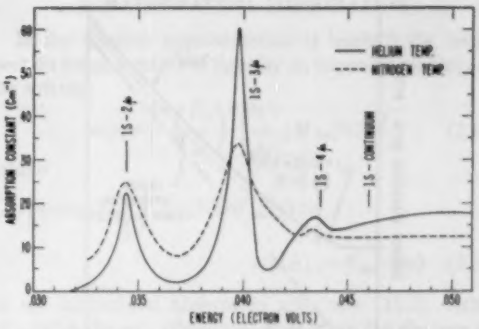


FIG. 2. Infrared absorption of a single crystal of silicon doped with boron [Burstein, Bell, and Henvis (unpublished)]. Further discussion of this data may be found in Burstein, Bell, Davisson, and Lax, J. Phys. Chem. **57**, 849 (1953).

The extent to which a given mode of vibration $q(\tau)$ of propagation constant τ participates in the broadening of a level depends on the extent to which a perturbation potential $q(\tau)\exp(i\tau\cdot r)$ shifts the electronic energy. To a first approximation, this shift may be computed by taking the mean of $\exp(i\tau\cdot r)$ over the state in question. It is clear that phase cancellations will occur unless the wavelength of the phonon is comparable to, or larger than the Bohr radius of the state. More precisely, the phonons of importance will be those which obey

$$\tau \lesssim 2/a \simeq 1.48\times 10^7 \text{ cm}^{-1}, \qquad (1.1)$$

when the Bohr radius a is given by:

$$a = (m/m_h)Ka_H = (m/m_h)K(\hbar^2/me^2) \simeq 13.5 \text{ A}. \qquad (1.2)$$

In (1.2) we have used a dielectric constant $K \simeq 11.5$. The effective mass for holes $m_h/m \simeq 0.45$ is determined by requiring that the hydrogenic formula yield agreement with the observed ground-state energy[7] $E_i \simeq 0.046 \pm 0.001$ ev. This "empirical" effective mass is intermediate between the two observed cyclotron resonance masses of the v_1 and v_2 bands (see Dexter and Lax).[6] It is hoped that the use of such an intermediate mass, and hence intermediate Bohr radius will partly cancel the error of neglecting the degeneracy in the bands.

The condition (1.1) applies to broadening of the $1s$ ground state. For broadening of the $2p$, $3p$, or $4p$ states, we should insert a Bohr radius two, three, or four times as large. Since the number of phonons with propagation constant $<2/a$ is proportional to $(2/a)^3$, the broadening of the $2p$, $3p$, and $4p$ states should be related to the broadening in the $1s$ state roughly by a factor $1/8$, $1/27$, and $1/64$ respectively. Thus, the broadening associated with a transition $1s-np$ is produced almost entirely by the broadening of the $1s$ level. For this reason, we expect that the $1s-2p$, $1s-3p$ and $1s-4p$ transitions show about the same width. This conclusion is in agreement with experiment (Fig. 2).

[7] Burstein, Bell, Davisson, and Lax, J. Phys. Chem. **57**, 849 (1953). This paper establishes the existence of excited impurity states which hitherto had only been conjectured. See this paper for a summary of the experimental procedure. Qualitatively similar results have also been obtained by R. Newman for In-doped silicon, Phys. Rev. **99**, 465 (1955).

[8] C. Kittel and A. H. Mitchell, Phys. Rev. **96**, 1488 (1954); M. A. Lampert, Phys. Rev. **97**, 352 (1955); W. Kohn and J. Luttinger, Phys. Rev. **97**, 1721 (1955); **98**, 915 (1955). Since this work was completed, wave functions for the acceptor states have also been obtained by W. Kohn (private communication).

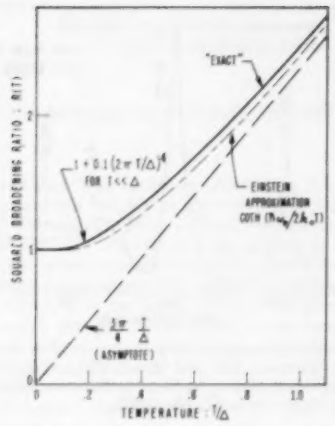


FIG. 3. The theoretical ratio, $R(T)$, of the squared broadening at the temperature T to its zero-point value, is plotted as a function of the dimensionless variable T/Δ. The characteristic temperature $\Delta=2\hbar v/(k_0 a)\simeq 94°$K in silicon, where v is the velocity of sound, a is the Bohr radius in the crystal and k_0 is Boltzmann's constant.

The modes of importance in the broadening can according to (1.1) be assigned to a characteristic temperature Δ according to

$$k_0\Delta\simeq\hbar\omega=\hbar v\tau\simeq\hbar v(2/a), \tag{1.3}$$

where k_0 is Boltzmann's constant, and v is the velocity of sound. For Silicon, we may use[9] $v=(C_{11}/\rho)^{\frac{1}{2}}\simeq 0.83\times 10^6$ cm/sec. This leads to a characteristic temperature $\Delta\simeq 94°$K. The ratio $R(T)$ of the mean squared broadening at temperature T to the zero-point broadening can then be approximated roughly[10] by the corresponding ratio of mean-squared oscillations of a harmonic oscillator of frequency $\omega=k_0\Delta/\hbar$:

$$R(T)\simeq\coth(\hbar\omega/2k_0T)=\coth(\Delta/2T). \tag{1.4}$$

Thus, a noticeable increase in broadening should occur at a temperature $T=\Delta$. In fact, $R(\Delta)$ is roughly 2, so that a 40 percent increase in the line width should be observed from absolute zero to 90°K. This is in good agreement with the actual increase in width between helium and nitrogen temperatures. The observed increase at low temperatures is strong evidence for the importance of long wavelength acoustic modes in the broadening; these temperatures are much smaller than the Debye temperature[11] $\theta\simeq 658°$K.

The good qualitative agreement between the experimental and theoretical temperature dependence of the broadening lends support to our proposed mechanism of broadening by the acoustical modes through the electron lattice interaction. A more stringent test would be a comparison between the predicted zero-point broadening and the experimental broadening at liquid helium temperatures. Our detailed theory, to be presented later, shows that the rms width of the line at 0°K can be written in the form [see Eq. (5.15)]:

$$\begin{aligned}\langle(\hbar\nu-\hbar\bar{\nu})^2\rangle^{\frac{1}{2}}&\simeq(\hbar/3\pi^2v\rho a^4)^{\frac{1}{2}}E_1\\&\simeq 2.3\times10^{-4}E_1,\end{aligned} \tag{1.5}$$

where $\rho\simeq 2.33$ g/cm^3 is the (x-ray) density of silicon, $v=0.83\times10^6$ is the sound velocity, and $a\simeq 13.5\times10^{-9}$ cm is the Bohr radius.

The deformation potential constant[12] E_1 measures the strength of the electron-lattice interaction and was chosen to be 15 ev to fit the experimental mobility at 300°K (see Sec. 4). Thus, the predicted rms width is about 3.6 millivolts, whereas the experimental full width at half-power is about 1 millivolt. (Experimental widths of donor states have a similar value.) For a Gaussian line shape, the full width at half-power is $(8\ln 2)^{\frac{1}{2}}\simeq 2.35$ times bigger than the rms width. This would increase the predicted width to about 8.4 mv. However, an investigation of the theoretical line shape shows that the line shape is far from Gaussian due to the presence of "motional narrowing." When this is taken into account roughly, [see Eq. (6.2)] the predicted line width becomes 3.0 mv or about three times the experimental width.

The discrepancy between theory and experiment is sufficiently small, considering the difficulties of making an absolute calculation, that the proposed broadening mechanism seems quite reasonable. It nevertheless may be of some interest to examine the possible reasons for the discrepancy. Perhaps the first assumption to be re-examined is the use of the Born-Oppenheimer[13] approximation. We have assumed that the electronic wave function and energy adjusts to the instantaneous positions of the nuclei. If the nuclei move too fast, this will not be the case: the energy levels will be less sensitive to the nuclear motion and a smaller broadening will result. We know, therefore, that our predicted broadening is an overestimate.

Usually the validity of the Born-Oppenheimer approximation is guaranteed by the large mass ratio[13]: in a molecule the electrostatic restoring forces that act on an electron and on a nucleus are comparable in size. The frequencies of oscillation of the electron and nucleus will then be proportional to the inverse square root of the respective masses. However, in a semiconductor, the impurity center electron is not responsible for the cohesion of the lattice and there is no necessary relation between the forces acting on this electron and the nuclei. The nuclear vibrations are determined by the usual cohesive forces which are practically independent of the presence of the impurity electron. The electron, on the

[9] The velocity of sound varies with orientation in silicon. The longitudinal velocity in the (1,0,0) direction is taken as representative.

[10] More precise estimates of the broadening ratio given by Eqs. (5.22), (5.25) and Fig. 3 agree closely with (1.4).

[11] N. Pearlman and P. H. Keesom, Phys. Rev. **88**, 398 (1952).

[12] J. Bardeen and W. Shockley, Phys. Rev. **80**, 72 (1950).

[13] M. Born and J. R. Oppenheimer, Ann. Physik **84**, 457 (1927); M. Born and K. Huang, *Dynamics of Crystal Lattices* (Oxford University Press, Oxford, 1954).

other hand, is subject to forces weaker than usual because it moves in a large orbit, subject primarily to the electrostatic attraction of the center reduced by the dielectric constant. The validity of the Born-Oppenheimer approximation must therefore be reexamined in detail for the case of boron doped silicon:

The root mean square velocity of a hole trapped in the ground state is $(2E_g/m_h)^{\frac{1}{2}} \simeq 1.9\times10^7$ cm/sec where the experimental ionization energy[7] is $E_g = 0.046 \pm 0.001$ ev, and the effective mass m_h is chosen to be $m_h = 0.45m$ [see Eq. (1.2), ff]. This may be compared with the velocity of sound in the material $v \simeq 0.83\times10^6$ cm/sec.[9] The rms velocity of a typical nucleus is of the order of $(k\theta/M_{Si})^{\frac{1}{2}} \simeq 10^5$ cm/sec, where θ is the Debye temperature[11] 650°K and the mass of a silicon atom is $M_{Si} = 4.65\times10^{-22}$ grams. Thus the electron moves considerably faster than the nuclei and will surely adjust somewhat to their motion.

However, the full validity of the Born-Oppenheimer approximation presumably requires the electron to traverse its orbit several times during one vibration of a nucleus. Thus, we should compare the frequencies of the two motions. The frequency of the electronic motion is of the order $E_g/h \simeq 7\times10^{13}$ cycles per second. The nuclear motion has a spectrum of frequencies whose maximum is of the order $k\theta/h \simeq 10^{14}$ cycles per second. If the broadening were predominantly produced by such high frequencies, the Born-Oppenheimer approximation would fail. However, as we remarked in Eq. (1.1) the phonons of importance have a propagation constant $\tau \leq 2/a \simeq 1.48\times10^7$ cm^{-1}. This result is only 0.13 of the maximum τ:

$$\tau_{max} = (6\pi^2 n_0)^{\frac{1}{3}} \simeq 1.14\times10^8 \text{ cm}^{-1} \tag{1.6}$$

obtained by converting one cell of the reciprocal lattice space to a sphere of equal volume with radius τ_{max}, (where $n_0 = 2.5\times10^{22}$ primitive cells per cm^3). The important vibrational frequencies will then be less than $(0.13)\times10^{14}$ cycles per second, so that the electron (whose frequency is 0.7×10^{14}) will traverse its orbit five times before the important normal coordinates traverse theirs once. From this, we may conclude that the Born-Oppenheimer approximation will yield a broadening that is of the right order of magnitude.

The major source of error that remains is the use of a one-band picture that neglects the degeneracy of the valence bands at the center of the Brillouin zone. (A further, related assumption is that the band energy near $k=0$ is approximately quadratic in k for energies of the order of k_0T.) This error could be remedied by recalculating V_{aa} using (3.1) with a more accurate wavefunction $\psi_a(r)$ for the impurity state. Such a change is not likely to reduce the theoretical answer by a factor as large as three. It seems much more likely that an error of a factor of three can be produced by attempting to determine the constant E_1 using the standard one-band formula for the mobility Eq. (4.6). These points will be discussed further in Secs. 4 and 6.

2. MATHEMATICAL FORMULATION

If the Condon approximation is made,[14] the cross section for absorption of light by an impurity center can be written:

$$\sigma(\nu) = \left[\frac{n}{\epsilon}\left(\frac{E_s}{E}\right)^2\right]\frac{8\pi^3\nu}{3c}|M_{ba}|^2 G(\nu), \tag{2.1}$$

where

$$G(\nu) = A\nu_m \sum_n \left|\int X_{bn}^*(x) X_{am}(x)\,dx\right|^2 \times \delta(E_{bn} - E_{am} - h\nu) \tag{2.2}$$

is the normalized absorption spectrum (15.2). Here $X_{am}(x)$ is the mth vibrational state when the electron is in state a and $X_{bn}(x)$ is the nth vibrational state when the electron is in state b, where x is an abbreviation for the 10^{23} nuclear coordinates or normal coordinates of the lattice, and m, n are abbreviations for the quantum numbers associated with the 10^{23} corresponding harmonic oscillators.

In the Born-Oppenheimer approximation, the effective Hamiltonian for nuclear vibration depends on the electronic state and will be denoted H_a or H_b for the ground or excited electronic states, respectively. Thus the eigenfunctions $X_{am}(x)$ and $X_{bn}(x)$ belong to two different Hamiltonians and are not orthogonal with respect to each other. As a consequence (2.2) permits multi-phonon transitions in which the individual normal modes of oscillation change their quantum numbers by zero or plus or minus one.

The difficulty in evaluating (2.2) lies in performing the sum over those final states n that satisfy the conservation of energy requirement. Huang and Rhys[9] were able to perform this sum by special devices that are applicable only when all the modes have the same frequency (e.g., an idealized optical branch). In I, we showed that the sum over final states could be performed by closure techniques after the energy conservation delta function was replaced by an integral representation and energy eigenvalues are replaced by the appropriate Hamiltonians H_a and H_b. Regardless of the nature of the vibration spectrum, we found

$$G(\nu) = h^{-1}\int_{-\infty}^{+\infty} \exp(-i2\pi\nu t)\,dt G(t), \tag{2.3}$$

where

$$G(t) = \text{av}[\exp(iH_b t/h)\exp(-iH_a t/h)] \tag{2.4}$$

still contains the average to be performed over initial states. The Boltzmann averaging symbol av acting on any operator O is an abbreviation for:

$$\text{av}(O) = \text{trace}[O\exp(-H_a/kT)]/\text{trace}[\exp(-H_a/kT)], \tag{2.5}$$

[14] See reference 1 for terminology and notation. For further discussion of (2.1) with reference to absolute absorption cross sections and broadening, see M. Lax, "The influence of lattice vibrations on electronic transitions in solids" in *Proceedings of the November 1954 ONR Conference on Photoconductivity* (John Wiley and Sons, Inc., New York, to be published).

so that (2.4) has been expressed in an invariant way in terms of the many-body Hamiltonians H_a and H_b.

A reduction of this many-body problem to a set of one-body problems can be performed if the normal coordinates q_j associated with H_a and H_b are the same, i.e., if H_a and H_b can be decomposed into terms depending on the same normal coordinates with no cross terms:

$$H_a=\sum_j H_{aj}(q_j,p_j),$$
$$H_b=\sum_j H_{bj}(q_j,p_j). \tag{2.6}$$

In this case, the complete average (2.4) can be factored into a product of averages each pertaining to an individual normal mode:

$$G(t)=\prod_j G_j(t), \tag{2.7}$$

$$G_j(t)=\mathrm{av}[\exp(iH_{bj}t/\hbar)\exp(-iH_{aj}t/\hbar)]. \tag{2.8}$$

This factorization was an essential simplification accomplished in I. The resulting averages (2.8) for individual harmonic oscillators are easy to perform by a wide variety of methods. An early derivation[15] was based on methods similar to Huang and Rhys. In I, ordered operator methods were used because of their elegance and simplicity. O'Rourke[5] pointed out the usefulness of Dirac density matrix methods, and included a quadratic term in the energy difference. Quadratic differences in H_b-H_a were not discussed in I because in general they will cause a mixing of the normal coordinates destroying the separability exhibited in (2.7). For the special case considered by O'Rourke[4] in which the quadratic energy difference contains only diagonal terms we can write:

$$\Delta E(q_j)=H_{bj}-H_{aj}=N^{-\frac{1}{2}}A_jq_j+N^{-1}B_jq_j^2. \tag{2.9}$$

Equations I(7.7), (7.10), and (8.6) then yield:

$$G_j(t)=\mathrm{av}\exp_{[-]}\left[i\int_0^t \Delta E(q_j(s))ds/\hbar\right] \tag{2.10}$$

$$=\exp\{N^{-1}C_j^2[i(\sin\omega_jt-\omega_jt)-(2\bar{n}_j+1)$$
$$\times(1-\cos\omega_jt)]+iN^{-1}(B_j/M\omega_j)(\bar{n}_j+\tfrac{1}{2})t\}, \tag{2.11}$$

with

$$C_j^2=A_j^2/(2\hbar M\omega_j^3). \tag{2.12}$$

To obtain $G(t)$ from (2.11) one need only sum over j in the exponent. The resulting exponent then agrees (if $B_j=0$) with the previous calculation for a linear energy shift I(8.12). The effect of the quadratic energy difference is to add a term in the exponent linear in t. Such a linear term merely shifts the origin of ν in (2.3), thus yielding the same frequency spectrum displaced by a constant amount.[16] The amount of the shift in energy units according to (2.3) and (2.11) is

$$\Delta(h\nu)=N^{-1}\sum_j(B_j\hbar/M\omega_j)(\bar{n}_j+\tfrac{1}{2}) \tag{2.13}$$

or

$$\Delta(h\nu)=\mathrm{av}[N^{-1}\textstyle\sum_j B_jq_j^2], \tag{2.14}$$

i.e., the amount of the energy shift may be obtained by replacing the quadratic part of the energy difference by its thermal average value. The change of this shift with temperature was used by O'Rourke (following a suggestion of Huang and Rhys) to explain the shift in the F-center absorption peak with temperature.[17] The constants B_j were not calculated, but $\sum B_j/N$ was chosen to give agreement with the experimental shift[17] (neglecting the fact that part of the shift may be ascribed simply to expansion of the lattice with increasing temperature).

The calculation made in (2.10) is easily extended to cubic and higher terms in the energy difference. However, a term of order $N^{-\frac{1}{2}}q_j^3$ will yield no contribution of order N^{-1}, and thus will not contribute to the result. Does this mean that anharmonic terms have no influence on the absorption spectrum? Surely a cubic term x^3 in the displacement of an impurity atom or a neighboring atom will affect the spectrum. Yet when x^3 is expressed in terms of normal modes the terms are of order $N^{-\frac{1}{2}}q_jq_kq_l$ that would be omitted in (2.9). This suggests that it may not be permissible to neglect cross terms among the quadratic terms. The diagonal quadratic terms will undoubtedly continue to yield an energy shift (2.14), but the cross terms will cause a mixing of modes. This mixing of modes may not change the peak or the width of the absorption spectrum appreciably, but it may scramble up or destroy fine structure that would otherwise be present.

In our calculation, we shall take account only of linear terms because there is no easy way to make a reliable estimate of the higher-order terms. The linear terms will be estimated from the electron-lattice interaction using the deformation potential approximation.[12,18] The electron-lattice interaction is weak in silicon and germanium —as shown by the high mobility. Thus the quadratic terms should be small. (As corroborating experimental evidence we note that the absorption line does not shift with temperature over a range in which appreciable broadening is observed.) We assume, therefore, that the influence of the quadratic terms on the first few moments of the spectrum are small—though they may have a significant effect on the higher moments, i.e., on details of the shape. We shall therefore use our calculation only to estimate the first few moments of the spectrum and not to make a detailed evaluation of the line shape.

3. EFFECTIVE VIBRATIONAL HAMILTONIANS

Let us suppose that the change in electronic potential associated with an arbitrary (small) displacement of the nuclei is $V(\mathbf{r},\cdots)$. The dots represent the dependence of

[15] M. Lax, Naval Research Laboratory Report 3973, June 17, 1952 (unpublished).

[16] If off-diagonal quadratic terms had been included, the shape of the spectrum would have changed as well. Further discussion of this point is given by Lax.[14]

[17] Further discussion of the F-center problem is given by Lax.[14]

[18] An analysis of the validity of the deformation potential approximation is being made by R. Rosenberg as part of a thesis at Syracuse University on absorption of light by free carriers.

this potential on the displacements of the nuclei, or the normal coordinates of the latter's motion. To our present accuracy, the dependence of V on these displacements is linear. If the trapped electronic wave function in a ground or excited state is $\psi_a(\mathbf{r})$ or $\psi_b(\mathbf{r})$ for the case of a perfect lattice, the change in energy due to the lattice vibration is V_{aa} or V_{bb} correct to terms linear in the displacements. And a typical matrix element important in thermal transition problems is V_{ba}, where

$$V_{ba}=[\psi_b(\mathbf{r}),V(\mathbf{r},\cdots)\psi_a(\mathbf{r})]$$

$$\equiv\int\psi_b^*(\mathbf{r})V(\mathbf{r},\cdots)\psi_a(\mathbf{r})d\mathbf{r}. \qquad (3.1)$$

However, neither the potential V nor the wave functions are known accurately unless a detailed knowledge of the band structure is available. We shall show nevertheless that an evaluation of (3.1) can be made without such information (for the nondegenerate case).

A wave function such as $\psi_a(\mathbf{r})$ can be expanded[19–21] in terms of Bloch orbitals $\psi(k,\mathbf{r})$ or Wannier orbitals $\alpha(\mathbf{r}-Q)$

$$\psi_a(\mathbf{r})=\sum b_a(k)\psi(k,\mathbf{r})=\sum c_a(Q)\alpha(\mathbf{r}-Q). \qquad (3.2)$$

If there is no degeneracy between bands, and if the trap states are sufficiently spread out, it is sufficient to expand in terms of Bloch waves or Wannier orbtals of one band[22,23] the conduction band for electrons, or the valence band for holes. (Actually the valence band in silicon contains three bands degenerate at $k=0$, one of which is split off by spin-orbit interaction.[18–20] Since a detailed analysis of the impurity states is not yet available[8] under such circumstances, we shall be content here to do the calculation as if only one band were present. Part of the error in such a procedure will be cancelled by using an effective mass determined from the ionization energy of boron impurities.)

Although we do not have detailed information concerning the Wannier orbitals in (3.2), we know that $c_a(Q)$ is a hydrogenic function of its discrete argument Q. We may regard $b_a(k)$ as the same wave function in the momentum representation:

$$b_a(k)=N^{-\frac{1}{2}}\sum c_a(Q)\exp(-i\mathbf{k}\cdot Q). \qquad (3.3)$$

If (3.2) and a corresponding equation for $\psi_b(\mathbf{r})$ are inserted into (3.1) we find:

$$V_{ba}=\sum_{kk'}b_b^*(k')b_a(k)V_{k'k}. \qquad (3.4)$$

In principle, (3.4) constitutes the evaluation of (3.1) since the matrix elements $V_{k'k}$ are the usual matrix elements of the theory of electrical conductivity. Inserting (3.3) into (3.4), we obtain a form

$$V_{ba}=\sum_{Q'Q}c_b^*(Q')c_a(Q)V_{Q'Q} \qquad (3.5)$$

expressed directly in terms of the hydrogenic functions $c(Q)$, with

$$V_{Q'Q}=N^{-1}\sum V_{k'k}\exp(i\mathbf{k}'Q'-i\mathbf{k}Q). \qquad (3.6)$$

The momentum wave functions $b(k)$ fall off rapidly for k larger than a reciprocal Bohr radius. Furthermore, this reciprocal Bohr radius is small compared to the size of the Brillouin zone—compare (2.3) and (2.4). Thus the dominant contribution to V_{ba} in (3.4) comes from $V_{k'k}$ with sufficiently small k and k' that the deformation potential[12] is likely to be a good approximation.[24] In the latter approximation $V_{k'k}$ is proportional to $(\mathbf{k}'-\mathbf{k})$. To effect a simplification of (3.5) and (3.6), however, all we need is that $V_{k'k}$ be a function only of $\mathbf{k}'-\mathbf{k}$, a result that remains valid even in a self-consistent theory of the electron-lattice interaction.[25] If we write $V_{k'k}=V(\Delta k,\cdots)$ where $\Delta k=k'-k$ and the dots are as usual the displacement or normal coordinates, then (3.5) becomes:

$$V_{ba}=\sum_\tau V(\tau,\cdots)N_{ba}(\tau), \qquad (3.7)$$

where

$$N_{ba}(\tau)=\sum_Q c_b^*(Q)\exp[i\tau\cdot Q]c_a(Q). \qquad (3.8)$$

Here τ is used as a summation index, replacing Δk. From (4.4), it can be seen to have the significance of a phonon propagation vector. The coefficients $c(Q)$ are normalized in the discrete sense:

$$\sum_Q|c_a(Q)|^2=1. \qquad (3.9)$$

Thus, $N_{aa}(\tau)$ is the discretized mean value of $\exp[i\tau\cdot Q]$ in the hydrogenic ground state. Since the ground state $c_a(Q)$ varies slowly as Q moves from one lattice point to another, we may replace the sum (3.8) by the corresponding normalized integral:

$$N_{aa}(\tau)=\int\exp(i\tau\cdot\mathbf{r})(\pi a^3)^{-1}\exp(-2r/a)d\mathbf{r}$$

$$=[1+(\tfrac{1}{2}a\tau)^2]^{-2}. \qquad (3.10)$$

For the first excited state, the $2p$ hydrogenic state, we have

$$N_{bb}(\tau)=\int\exp(i\tau\cdot\mathbf{r})(32\pi a^5)^{-1}[r\cos(r,e)]^2\exp(-r/a)d\mathbf{r}$$

$$=[1+(a\tau)^2]^{-4}[1-(a\tau)^2-4(a\tau)P_2(\cos(\tau,e))],$$

when we have chosen a state with $l=1$, $m=0$ with our z axis in the direction of the polarization vector e of the light quantum. This is the state that will absorb light polarized in the z-direction.

[18] G. H. Wannier, Phys. Rev. **52**, 191 (1937).
[19] J. C. Slater, Phys. Rev. **76**, 1592 (1949); G. F. Koster and J. C. Slater, Phys. Rev. **95**, 1167 (1954).
[20] H. M. James, Phys. Rev. **76**, 1602, 1611 (1949).
[21] E. N. Adams, Phys. Rev. **85**, 41 (1952), J. Chem. Phys. **21**, 2013 (1953).
[22] P. Feuer, Phys. Rev. **88**, 92 (1952).
[24] Strictly speaking, one should use a generalization of the deformation potential to include scattering by transverse modes. See, however, reference 26. For degenerate bands one should also include the effect of interband transitions.
[25] J. Bardeen, Phys. Rev. **52**, 688 (1937).

4. DEFORMATION POTENTIAL APPROXIMATION

A comparison of (3.7) and (3.10) shows that the evaluation of V_{aa} requires an accurate knowledge of $V(\tau,\cdots)$ only for $\tau<(2/a)$, i.e., we are concerned with phonons of sufficiently long wavelength compared to the lattice spacing [compare (1.1) and (1.6)] that the deformation potential approximation should be adequate.[24] We shall in this paper neglect entirely the contribution of the optical modes and assume that for acoustical modes the two particles in the primitive cell of the diamond structure move in unison. (An analysis of the errors involved in the deformation potential approximation is in progress.[18]) Thus we can simplify our notation by proceeding as if we were dealing with a lattice containing one atom per cell whose mass M is twice that of a silicon atom M_{Si}. An arbitrary deformation $\delta\mathbf{R}(\mathbf{r}_j)$ can then be expressed in terms of the complex normal coordinates $q_a(\tau)$ by

$$\delta\mathbf{R}(\mathbf{r}_j)=N^{-\frac{1}{2}}\sum_\tau\sum_{a=1}^{3}\mathbf{e}_a(\tau)q_a(\tau)e^{i\tau\cdot r_j}, \tag{4.1}$$

where $\mathbf{e}_a(\tau)$ are the unit polarization vectors, and for sufficiently small τ we may adopt the convention that $\mathbf{e}_3(\tau)=\tau/\tau$ is the longitudinal mode. The deformation potential is then given by[26]

$$V(\mathbf{r},\cdots)=E_1\,\mathrm{div}\delta\mathbf{R}(\mathbf{r})=iE_1N^{-\frac{1}{2}}\sum_\tau\tau q_3(\tau)e^{i\tau\cdot r} \tag{4.2}$$

and the standard matrix element is given by

$$\begin{aligned}V(\Delta k,\cdots)&=[\psi(\mathbf{k}',\mathbf{r}),V(\mathbf{r},\cdots)\psi(\mathbf{k},\mathbf{r})]\\&\simeq iE_1N^{-\frac{1}{2}}\Delta k q_3(\Delta\mathbf{k}),\end{aligned} \tag{4.3}$$

which depends on the approximation

$$[\psi(\mathbf{k}',\mathbf{r}),e^{i\tau\cdot r}\psi(\mathbf{k},\mathbf{r})]\simeq\delta_{\Delta k\tau}. \tag{4.4}$$

whose validity follows from the smallness of Δk.

Since only longitudinal modes are involved hereafter, we shall drop the subscript 3 on q_3. Furthermore, $q(\tau)$ can be written

$$q(\tau)=[\hbar/2M\omega(\tau)]^{\frac{1}{2}}[a(\tau)-a^+(\tau)] \tag{4.5}$$

where $a(\tau)$ and $a^+(\tau)$ are conventional destruction and creation operators for phonons of propagation constant τ with matrix elements $[n(\tau)]^{\frac{1}{2}}$ and $[n(\tau)+1]^{\frac{1}{2}}$ respectively.

The matrix elements (4.3) and (4.5) lead to the customary formula for the mobility,[12]

$$\mu=\left(\frac{8\pi}{9}\right)^{\frac{1}{2}}\frac{e\hbar^4n_0Mv_s^2}{E_1^2(kT)^{\frac{1}{2}}(m^*)^{5/2}}, \tag{4.6}$$

for the case of spherical energy surfaces.[27]

An estimate of E_1 for holes in silicon can be obtained from (4.6). Using $\mu_p\simeq500$ cm^2/volt-sec[28] at 300°K, a velocity of sound $v\simeq0.83\times10^6$ cm/sec, an effective mass $m^*\simeq0.45m$, and $M=2M_{Si}$, we find that $E_1\simeq15$ ev. Our estimate is based on a high-temperature mobility because the broadening mechanism automatically selects lattice scattering in preference to impurity or dislocation scattering: The trapped hole will only be scattered by vibrations or imperfections in its immediate vicinity. Imperfections sufficiently close to the impurity atom to be of importance will very likely shift the entire line by more than its width in a somewhat random way and thus will not appear with any appreciable intensity in the observed absorption spectrum.

Thus, it is necessary to determine separately the effect of lattice scattering, which is partially accomplished by using a room temperature mobility. However, the experimental mobility[28] $\mu_p\simeq2.4\times10^8\,T^{-2.3}$ disagrees with the $T^{-1.5}$ power predicted by the deformation theory. If, for example, we were to fit our constant E_1 at 100°K instead of 300°K, we would find $E_1\simeq10$ ev. This leads to a reduction of a factor of 2 in the width and leads to much better agreement with experiment. We shall prefer, however, for the sake of conservatism to use the room temperature fit, and regard the factor of 2 as an uncertainty in the predicted broadening because of the absence of an adequate mobility theory.

5. CHARACTERISTIC FUNCTION AND THE MOMENTS

Combining (3.7) and (4.3), the difference between the two Hamiltonians H_b-H_a (aside from a constant term) is given by

$$V_{bb}-V_{aa}=N^{-\frac{1}{2}}\sum A(\tau)q(\tau), \tag{5.1}$$

where

$$A(\tau)=iE_1\tau N(\tau);\quad N(\tau)=N_{bb}(\tau)-N_{aa}(\tau). \tag{5.2}$$

The characteristic function $G(t)$ can now be obtained from (2.11) or (I8.12) providing we merely replace A_j^2 by $|A(\tau)|^2$ since the normal coordinates $q(\tau)$ are complex (see Appendix 1).

$$G(t)=\exp[i2\pi\nu_0t+f(t)], \tag{5.3}$$

$$\begin{aligned}f(t)=\langle|C(\tau)|^2[&i\sin\omega_\tau t\\&-(2\bar{n}_\tau+1)(1-\cos\omega_\tau t)]\rangle,\end{aligned} \tag{5.4}$$

$$|C(\tau)|^2=E_1^2\tau^2|N(\tau)|^2/(2\hbar M\omega_\tau^3). \tag{5.5}$$

The average, $\langle\ \rangle$ taken in (5.4) is defined as in (I5.21) to be an average over the modes τ, i.e., N^{-1} times the sum

[24] See reference 12, Eq. (A30). The neglect of the transverse contribution is not serious here, as pointed out by Bardeen and Shockley. The contribution of transverse modes can be included roughly, providing we understand that E_1^2 now represents the sum of the squares of three deformation constants. Since E_1^2 will later be chosen by comparison with experiment, transverse contributions are automatically included. We shall, therefore, use the original deformation potential thereby avoiding complicated notation. The most serious error made by this procedure is that we do not explicitly take account of interband transitions.

[27] For ellipsoidal surfaces centered at $k=0$ with effective masses m_1, m_2, m_3, Eq. (4.6) is still valid providing m^* is defined by:

$$(m^*)^{5/2}=3(m_1m_2m_3)^{\frac{1}{2}}/(m_1^{-1}+m_2^{-1}+m_3^{-1}).$$

For an eight-ellipsoid picture such as prevails in germanium,[6] (4.6) must be reduced by the presence of interellipsoid transitions. The latter effect can be shown to be small in germanium however.[18]

[28] M. B. Prince, Phys. Rev. **93**, 1204 (1954).

of the contribution of the N different modes τ. The energy $h\nu_0$ constitutes the thermal activation energy,[1] and will be defined more precisely in (5.40).

The moments of a distribution function can be obtained directly from the derivatives of the corresponding characteristic function at the origin, see I (4.2), I (8.14). Explicit formulas for the moments are identical to I (8.16)

$$h\bar{\nu}=h\nu_0\pm\langle|C(\tau)|^2\hbar\omega_\tau\rangle, \tag{5.6}$$

$$\langle(h\nu-h\bar{\nu})^2\rangle=\langle|C(\tau)|^2(2\bar{n}_\tau+1)(\hbar\omega_\tau)^2\rangle, \tag{5.7}$$

$$\langle(h\nu-h\bar{\nu})^3\rangle=\pm\langle|C(\tau)|^2(\hbar\omega_\tau)^3\rangle, \tag{5.8}$$

$$\langle(h\nu-h\bar{\nu})^4\rangle=3|\langle(h\nu-h\bar{\nu})^2\rangle|^2 +\langle|C(\tau)|^2(2\bar{n}_\tau+1)(\hbar\omega_\tau)^4\rangle. \tag{5.9}$$

The upper signs refer to absorption, the lower to emission. The first moment or mean absorption energy $h\bar{\nu}$ will be discussed later in this section.

The second moment (5.7) is a measure of the breadth of the absorption line in the transition $a\rightarrow b$. The breadths of the individual levels may be obtained by using N_{aa} or N_{bb} in place of $N(\tau)$ in (5.5). The squared breadth associated with the transition, however, may differ from the sum of breadths of the individual levels through a cross or correlation term:

$$\mathrm{av}(V_{bb}-V_{aa})^2 =\mathrm{av}V_{aa}^2+\mathrm{av}V_{bb}^2-2\,\mathrm{av}(V_{aa}V_{bb}). \tag{5.10}$$

The significance of (5.10) may be visualized as follows: during a lattice displacement the energy of level a changes by an amount V_{aa}. Thus, $\mathrm{av}V_{aa}^2$ is the mean-squared excursion of width of state a. Similar remarks apply to state b. However, the width of an absorption line is determined by $V_{bb}-V_{aa}$, the shift of line b relative to a, so that (5.10) automatically takes account any correlation in the shifts of the two lines.

We shall now calculate the broadening of the ground hydrogen-like state, by inserting (3.10) and (5.5) into (5.7):

$$\langle(h\nu-h\bar{\nu})^2\rangle_{aa}=\frac{hE_1^2}{Mv_sa}\left\langle\left[\frac{x}{(1+x^2)^4}(1+2\bar{n}_x)\right]\right\rangle, \tag{5.11}$$

where

$$x=\tfrac{1}{2}a\tau, \quad \bar{x}=\tfrac{1}{2}a\tau_{\max}\simeq 7.7, \tag{5.12}$$

$$\langle[\ ]\rangle=(3/\bar{x}^3)\int_0^{\bar{x}}[\ ]dx \tag{5.13}$$

and

$$\bar{n}_x=[\exp(x\Delta/T)-1], \tag{5.14}$$

with

$$\Delta=(2\hbar v/k_0a)\simeq 94^\circ\mathrm{K}.$$

The zero-point broadening, obtained by setting $\bar{n}_x=0$ in (5.11) may be calculated readily by noting that

$$\int_0^\infty x^2(1+x^2)^{-4}dx=(1/12); \quad [\hbar/(Mva)]\simeq 10^{-4}, \tag{5.4}$$

so that

$$[\langle(h\nu-h\bar{\nu})^2\rangle_{aa}]^{\frac{1}{2}}\simeq\left(\frac{\hbar}{4Mva\bar{x}^3}\right)^{\frac{1}{2}}E_1 \simeq 2.3\times10^{-4}E_1\simeq 3.6\times10^{-3}\ \mathrm{ev}. \tag{5.15}$$

The ratio of the broadening in the $2p$ state to that in the $1s$ state is given by

$$\frac{\langle V_{bb}^2\rangle}{\langle V_{aa}^2\rangle}=\frac{\langle x(1+4x^2)^{-6}[1-4x^2-8xP_2(\cos(x,e))]^2\rangle}{\langle x(1+x^2)^{-4}\rangle} \tag{5.16}$$

$$=\frac{\langle x(1+4x^2)^{-6}[1+(24/5)x^2+16x^4]\rangle}{\langle x(1+x^2)^{-4}\rangle}, \tag{5.17}$$

$$\frac{\langle V_{bb}^2\rangle}{\langle V_{aa}^2\rangle}=89/(5600)\simeq 0.016, \tag{5.18}$$

where the step from (5.15) to (5.16) involved an average over the orientation of the light polarization vector $\mathbf{e}$ relative to the propagation direction.

We may conclude from (5.18) that the broadening of the $2p$ state is only one-eighth that of the ground state. The physical reason for this is that the Bohr radius of a $2p$ state is twice that of a $1s$ state and the number of phonons with sufficiently large wavelength to produce broadening is correspondingly reduced. The broadening of the $3p$ and $4p$ states will be even less. The entire width of the $1s-2p$, $1s-3p$ and $1s-4p$ transitions can therefore be ascribed to the ground state. (There is, of course, also broadening of the $1s$-continuum transition, but this can only be observed as a rounding of the edge).

The ratio of the second moment at temperature T to its zero-point value is given by

$$R(T)=1+24\int_0^\infty x^2(1+x^2)^{-4}[\exp(x\Delta/T)-1]^{-1}dx \tag{5.19}$$

for the ground state. At sufficiently high temperatures, the second moment becomes linear in T:

$$R(T)\underset{T\rightarrow\infty}{\rightarrow}(3\pi/4)(T/\Delta), \tag{5.20}$$

whereas at low temperatures it approaches its zero-point value as T^4:

$$R(T)\underset{T\rightarrow 0}{\rightarrow}1+(8\pi^4/5)(T/\Delta)^4. \tag{5.21}$$

An approximate expression for $R(T)$ may be obtained by replacing the distribution of modes by a single effective frequency. In this case

$$R(T)\approx\coth(\tfrac{1}{2}\hbar\omega_e/k_0T), \tag{5.22}$$

where the best choice for the effective frequency,[14] using Eq. I (5.22) and I (5.23), is

$$\hbar\omega_e=(8/3\pi)k_0\Delta. \tag{5.24}$$

The Einstein approximation (5.22) approaches unity exponentially at low temperatures in contrast to the power law behavior of (5.21) which follows from the more accurate Debye treatment.

The integral (5.19) can be carried out exactly using contour integration methods with the result:

$$R(T)=(1/z)[(3/8)+(3/4)z^2\psi'(z) -(3/4)z^3\psi''(z)-(1/4)z^4\psi'''(z)], \quad (5.25)$$

where

$$z=\Delta/(2\pi T) \quad (5.26)$$

and

$$\psi(z)=d/dz\ln\Gamma(1+z);\quad \psi'(z)=\sum_{\nu=1}^{\infty}\frac{1}{(z+\nu)^2}. \quad (5.27)$$

The digamma, trigamma, tetragamma, and pentagamma functions $\psi(z)$, $\psi'(z)$, $\psi''(z)$ and $\psi'''(z)$ are tabulated by Davis.[29]

A comparison of the "exact" broadening (5.26) and the "Einstein" broadening *versus* temperature is shown in Fig. 3.

Equations (5.19)–(5.21) describing the change in broadening with temperature apply to trapped electrons as well as holes. However, it must be remembered that the temperature Δ in (5.13) varies inversely with the Bohr radius a; $\Delta_e/\Delta_h=a_h/a_e=m_e/m_h$ where the last equality applies if both states are hydrogenlike.

The ratio of the zero-point broadening for electrons to that for holes according to (5.15) and (4.6) is given by:

$$\frac{\langle(h\nu-h\bar{\nu})^3\rangle_e}{\langle(h\nu-h\bar{\nu})^3\rangle_h}\underset{T\to 0}{\to}\left(\frac{E_e m_e^2}{E_h m_h^2}\right)^2=\left(\frac{m_e}{m_h}\right)^{\frac{1}{2}}\frac{\mu_h}{\mu_e}. \quad (5.28)$$

The third moment according to (5.8) is independent of temperature and can be written in the form:

$$\langle(h\nu-h\bar{\nu})^3\rangle\simeq\pm\frac{E_1^2\hbar^2}{Ma^2}\frac{6}{\bar{x}^3}\int_0^\infty\frac{x^4}{(1+x^2)^4}dx \simeq\frac{3\pi}{16}\frac{E_1^2\hbar^2}{Ma^2\bar{x}^3}. \quad (5.29)$$

The contribution of the second term in (5.9) to the fourth moment can be written

$$\langle|C(\tau)|^2(2\bar{n}_\tau+1)(\hbar\omega_\tau)^4\rangle\equiv\frac{2E_1^2\hbar^2v_s}{Ma^2\bar{x}^3}R_4(T), \quad (5.30)$$

where

$$R_4(T)\simeq 6\int_0^\infty\frac{x^3}{(1+x^2)^4}\coth\left(\frac{x\Delta}{2T}\right)dx \quad (5.31)$$

reduces to unity at absolute zero. In fact at sufficiently low temperatures ($T\ll\Delta$)

$$R_4(T)\to 1+1440\zeta(5)(T/\Delta)^6, \quad (5.32)$$

where $\zeta(5)=1.0369$, and at sufficiently high temperatures ($T\gg\Delta$)

$$R_4(T)\to 3\pi T/(8\Delta) \quad (5.33)$$

whereas at intermediate temperatures $T\sim\Delta$, or higher, a sufficiently accurate expression can be obtained by a mean value estimate of (5.31):

$$R_4(T)\simeq\coth[8\Delta/(3\pi T)]. \quad (5.34)$$

A dimensionless description of the shape of the absorption curve can be obtained from the skewness and kurtosis:

$$\gamma_3(T)=\frac{\langle(h\nu-h\bar{\nu})^3\rangle}{[\langle(h\nu-h\bar{\nu})^2\rangle]^{\frac{3}{2}}} =\frac{3\pi}{2}(\bar{x})^{\frac{1}{2}}\frac{Mv_s^2}{E_1}\left(\frac{\hbar}{Mv_sa}\right)^{\frac{1}{2}}[R(T)]^{-\frac{3}{2}}. \quad (5.35)$$

$$\gamma_4(T)=\frac{\langle(h\nu-h\bar{\nu})^4\rangle}{[\langle(h\nu-h\bar{\nu})^2\rangle]^2}-3 =32(\bar{x})^3\left(\frac{Mv_s^2}{E_1}\right)^2\frac{\hbar}{Mv_sa}\frac{R_4(T)}{[R(T)]^2}. \quad (5.36)$$

$$\gamma_3(0)\simeq 1.8;\quad \gamma_4(0)\simeq 8.9. \quad (5.37)$$

From (5.37) we see that the skewness and kurtosis are quite appreciable near absolute zero. However, they decrease at higher temperatures. Since $\gamma_3(T)$ semiclassical $=0$, and $\gamma_4(T)$ semiclassical $=0$, we see that the semiclassical approximation is not valid at low or even moderate temperatures.

It may be of interest to note that

$$\langle|C(\tau)|^2\rangle=E_1^2a/(8\hbar Mv_s^3\bar{x}^3) \quad (5.38)$$

$$\simeq E_1^2/590\simeq 0.38, \quad (5.39)$$

where E_1 has been set equal to 15 ev in (5.39). The smallness of this parameter indicates the feasibility of expanding $\exp[f(t)]$ in powers of $f(t)$ in (5.3), i.e., of expressing the results in terms of zero-phonon, one-phonon, two-phonon contributions, etc. (The parameter $\langle|C(\tau)|^2\rangle$ corresponds to the S of Huang and Rhys and indicates the average number of phonons to be expected in a transition.) This point will be discussed again in Sec. 6.

In I, we showed that because of the interaction between the electron and the lattice, the lattice relaxes to a new set of equilibrium positions one for state a, and one for state b. These positions are the minima of the parabolas that give the configurational energy of the system (see Fig. 1 of I). The energy difference between these minima denoted $h\nu_0$, corresponds to the thermal activation energy between the states a and b. We

[29] H. T. Davis, *Tables of the Higher Mathematical Functions* (Principia Press, Bloomington, 1933), Vols. I and II.

showed in (I 6.7) that

$$h\nu_0 = E_b - E_a + \langle [(A_j^a)^2 - (A_j^b)^2]/(2M\omega_j^2) \rangle, \quad (5.40)$$

where $E_b - E_a$ is the electronic energy difference with the nuclei constrained to the perfect lattice positions. The constants A_j^a and A_j^b are the linear interaction constants appropriate to state a or b respectively [see (5.2)]. For our purposes we have shown that the interaction in the $2p$ state is negligible compared to that in the $1s$ state, so that we may set $A_j^b = 0$, $A_j^a = A$. For the present case, then,

$$h\nu_0 = E_b - E_a + \langle |C(\tau)|^2 \hbar\omega \rangle \quad (5.41)$$

and

$$h\bar{\nu}_a = E_b - E_a + 2\langle |C(\tau)|^2 \hbar\omega \rangle \quad (5.42)$$

follows from (5.6). Thus the absorption energy $h\bar{\nu}_a$ will shift from the constrained lattice value. An estimate of this shift may be obtained by noting that

$$\langle \hbar\omega |C(\tau)|^2 \rangle = \frac{3\pi}{8} \frac{\langle (h\nu - h\bar{\nu})^2 \rangle_{T=0}}{k_0 \Delta}$$

$$\simeq \frac{3\pi}{8} \frac{(3.6 \text{ mv})^2}{8.1 \text{ mv}} \simeq 1.9 \text{ mv}. \quad (5.43)$$

Thus a total shift of 3.8 mv could occur. Since our theoretical breadth is too large by a factor three, the shift is more likely about 1.3 mv. This shift cannot, of course, be observed directly because it is temperature independent. It could affect slightly, a comparison between theoretical and experimental estimates of impurity level energies.

6. COMPARISON WITH EXPERIMENT

There are two qualitative experimental results that bear emphasis and explanation: Only a finite number (three) of transitions has been observed, and the width of the absorption line is approximately the same in each case. The second result we have already explained by noting that the broadening in the three transitions $1s-2p$, $1s-3p$, $1s-4p$ is primarily the broadening of the $1s$ level. The $2p-3p-4p$ levels have a much smaller broadening because their Bohr radii are larger, and the number of lattice modes with sufficiently long wavelength to produce broadening is correspondingly reduced.

It has been previously suggested that only a finite number of excited states exist because the wave functions associated with different impurity centers will eventually begin to overlap.[30] This is a limiting factor only at very high impurity concentrations, however. The crystal for which the measurements in Figs. 1 and 2 were taken (referred to as Si 155 in reference 7) is a fairly pure specimen having about 5×10^{15} free holes at room temperature. Thus the separation between boron centers is of the order of 1000 A. This is enough larger than the first Bohr radius 13.5 A that the centers may be regarded as independent.

A more elementary explanation can be given for the fact that only a finite number of transitions is observed—namely that such transitions would not be resolved since (a) they are close together in energy: they must all occur between 0.043 ev and 0.046 ev; see Fig. 2; (b) they should have widths of the order of 0.001 ev equal to those for the $1s-2p$, $1s-3p$, or $1s-4p$ transition; (c) they should have low intensities.

Another striking experimental result is the appreciable increase in width (about 40 percent) between liquid helium and liquid nitrogen temperatures. This is evidence that the phonons of importance have low energies $\hbar\omega \simeq k_0(80)$. Our analysis agrees with this by showing that the phonons of importance have wavelengths about equal to π times a Bohr radius. These phonons have a characteristic temperature of $\Delta = 2\hbar v/k_0 a \simeq 94°\text{K}$ in silicon, see (5.14). The squared broadening ratio is plotted *versus* T/Δ in Fig. 3. At $T = \Delta$, the squared broadening ratio is slightly over 2 in qualitative agreement with the experimental 40 percent increase.

Although the qualitative agreement with experiment suggests that our picture of the broadening as due to interaction with long wavelength acoustic modes is correct, our absolute estimate of the line width is slightly too large. The experimental full width at half power (F.W. at H.P.) is about one millivolt. The rms width [Eq. (5.15)] is about 3.6 millivolts. To convert from the rms width to the full width at half power, we need to know the shape of the line. For a Gaussian shape the

$$\text{F.W. at H.P.} = (8 \ln 2)^{\frac{1}{2}} [\langle (h\nu - h\bar{\nu})^2 \rangle]^{\frac{1}{2}} \simeq (2.35)(3.6) \simeq 8.4 \text{ mv} \quad (6.1)$$

at absolute zero (or liquid helium temperatures).

However, the line shape at absolute zero may be expected to be far from Gaussian as an examination of the kurtosis (5.36) shows. A large kurtosis implies a large fourth moment for the same second moment—so that more of the strength of the line is in the wings—with a correspondingly narrower, more peaked center. This "motional narrowing" of the line is a quantum mechanical effect that would not have appeared in a semiclassical treatment and results from the weakness of the electron-lattice interaction in silicon. A rough estimate of the importance of the line shape on the width may be obtained by comparing with the case of exchange narrowing in paramagnetic resonance treated by Anderson and Weiss.[31] Their estimate of the F.W. at H.P. $\simeq 2\omega_p^2/\omega_e$ can be rewritten in the form

$$\text{F.W. at H.P.} \simeq (2\pi/\gamma_4)^{\frac{1}{2}} [\langle (h\nu - h\bar{\nu})^2 \rangle]^{\frac{1}{2}} \simeq (0.84)(3.6) \simeq 3.0 \text{ mv} \quad (6.2)$$

by using the fourth moment to estimate ω_e. Because of

[30] K. S. Shifrin, J. Tech. Physics (U.S.S.R.) **14**, 43 (1944).

[31] P. W. Anderson and P. R. Weiss, Revs. Modern Phys. **25**, 269 (1953).

the large kurtosis, an appreciable narrowing results, and the estimated width of 3 millivolts is only about three times too large.

One of the possible reasons that the theoretical answer is too large is that the calculation was made using the Born-Oppenheimer approximation which inevitably overestimates somewhat the sensitivity of the electronic energy to the nuclear coordinates. This point was discussed in detail in Sec. 1. Another reason for error in our calculation of the electronic energies was the use of a simplified one band picture rather than the use of a degenerate band picture for the valence band in silicon. This error should be partly cancelled by using an effective mass determined from the experimental binding energy. Another approximation was the use of the deformation potential which neglects any contribution of transverse modes to the scattering matrix elements. This should be partly cancelled by the use of the experimental mobility and the deformation potential theory in the determination of the electron-lattice constant. Perhaps the worst error is that constant E_1 is estimated neglecting the possibility of hole scatterings between the degenerate bands and quartic terms in the energy-momentum relationship. Because of this neglect, we have probably overestimated E_1 and hence overestimated our theoretical width. (If we had used the 100°K mobility to determine E_1, the predicted broadening would have been reduced by a factor of $3^{\frac{1}{2}}\simeq 2.35$).

As a concluding remark, we would like to explain why we used the Anderson-Weiss narrowing estimate (6.2) rather than working directly from the line shape which is presumably known from (2.3) and (5.3). Because of the smallness of the parameter E_1 in silicon $\langle|C(\tau)|^2\rangle \simeq (E_1/24)^2 \simeq 0.4$ and it becomes legitimate to expand (5.3) in powers of $f(t)$. The term of order $[f(t)]^n$ then corresponds to an n-phonon process. The zeroth-order term in this expansion corresponds to a delta function, i.e., an infinitely sharp line at the electronic energy difference. The one- and two-phonon processes, etc., yield a continuous background. Because the integrated intensity is independent of E_1, a large E_1 (the semiclassical limit) would throw most of the intensity into the many phonon processes with a corresponding reduction in the intensity of the center (zero-phonon) line. It is then more legitimate to approximate the line shape by a continuous curve. For small E_1, as is the case in silicon, a fair fraction of the total intensity resides in the center line and no simple description of the line shape can be given. Of course, an infinitely sharp center line could never be observed because of lack of perfect resolution in the detection instruments. The fact that our line width increases from helium to nitrogen temperatures, however, shows that the observed broadening cannot be entirely instrumental.

One possible theoretical explanation is that it is not possible to neglect the terms in the energy difference of second order in the nuclear displacements even when such terms are small. For such terms imply that the normal coordinates of the lattice are different in the ground and excited electronic states. There is then, strictly speaking, no such thing as a zero-phonon process. The presence of a small second-order term could then conceivably result in a continuous line shape without appreciably affecting the first few moments of the spectral distribution. Thus, we have assumed that our calculated values of the first few moments are more reliable than our calculated line shape and have used them to make a crude estimate of the line width.

APPENDIX: INTRODUCTION OF REAL COORDINATES

The complex linear energy difference (5.1) can be rewritten in the symmetrized form:

$$\Delta E(q)=N^{-\frac{1}{2}}\sum_{\tau>0}[A(\tau)q(\tau)+A(-\tau)q(-\tau)], \quad \text{(A1)}$$

where we ignore the term $\tau=0$ since $A(0)=0$, and in any case, one term will make no difference after the limit $N\to\infty$ is taken. Since ΔE is a Hermitian operator and $q(-\tau)=\pm q^+(\tau)$ we must have also $A(-\tau)=\pm A^*(\tau)$. If we introduce these relations into (A1) and define

$$\begin{aligned} A(\tau)&=[A_1(\tau)+iA_2(\tau)]/\sqrt{2},\\ q(\tau)&=[q_1(\tau)+iq_2(\tau)]/\sqrt{2}, \end{aligned} \quad \text{(A2)}$$

then

$$\Delta E(q)=\sum_{\tau>0}[A_1(\tau)q_1(\tau)+A_2(\tau)q_2(\tau)]. \quad \text{(A3)}$$

Since the two modes $q_1(\tau)$ and $q_2(\tau)$ have the same frequency $\omega(\tau)$, any sum of the form

$$\begin{aligned} f&=\sum_{\tau>0,\,\alpha=1,2} A_\alpha^2(\tau)h(\omega(\tau))\\ &=\sum_{\tau>0}[A_1^2(\tau)+A_2^2(\tau)]h(\omega(\tau)) \end{aligned} \quad \text{(A4)}$$

can be rewritten using the evenness of $\omega(\tau)$ and $|A(\tau)|^2$ in the form

$$f=\sum_{\text{all }\tau}|A(\tau)|^2h(\omega(\tau)), \quad \text{(A5)}$$

where $h(\omega)$ is an arbitrary function of ω. The function $f(t)$ defined by (5.3) and (2.11) is a sum of this form. Thus, we may apply the formulas of reference 1 based on the real sum $\sum A_jq_j$ by replacing A_j^2 by $|A(\tau)|^2$.

Criterion for the Occurrence of Luminescence*

D. L. DEXTER, *Institute of Optics, University of Rochester, Rochester, New York and United States Naval Research Laboratory, Washington, D. C.*

AND

C. C. KLICK AND G. A. RUSSELL, *United States Naval Research Laboratory, Washington, D. C.*

(Received July 5, 1955)

A mechanism is suggested for the nonradiative quenching of excited states of simple impurity centers, namely a crossover to the ground electronic state of the center immediately after excitation and before thermal equilibrium has had time to come about. It is proposed that this mechanism is responsible for the absence of strong luminescence in some impurity and color centers, e.g., the *F*-center in alkali-halide crystals. The quenching process proposed here is applicable only for certain geometries of the configuration coordinate diagram, which are discussed with a view to a criterion for strong luminescence. The quenching mechanism is consistent with known configuration coordinate curves, and leads to a prediction which is experimentally verified, to the effect that the wavelength of emitted radiation is not more than twice that of the light absorbed by the center. The same mechanism shows the possibility of photoconductivity arising from absorption in discrete lines even at low temperatures.

IN a simple impurity center the absorption of light in the center raises it to a bound excited state from which it returns to the ground state either by a radiationless transition or by a luminescent transition. Although a large number of luminescent and nonluminescent systems has been investigated experimentally, the properties which determine whether or not a particular system luminesces have not been well understood.

This problem has become of special interest recently in connection with the luminescence of the *F*-center in alkali halides. Huang and Rhys[1] have predicted that a multiphonon radiationless transition in the *F*-center would have a negligible probability at low temperatures and Pekar[2] has indicated that a variety of alkali halides should have their *F*-center emission near 10 000 A. However, an experimental investigation by Klick[3] from 2500 to 25 000 A indicated no emission from the *F*-center with a quantum efficiency larger than three percent over most of this range. Botden, Van Doorn, and Haven[4] have recently reported emission from colored alkali halides quenched from elevated temperatures to liquid nitrogen temperature, but the efficiency of this emission is reported to be only of the order of one percent. These authors tentatively suggest that this emission arises from the *F*-center. In any case the large majority of *F*-centers return to their ground state by processes which do not involve radiation.

Markham[5] and Meyer[6] have suggested that more refined calculations on the multiphonon transitions could show that these transition probabilities were underestimated by Huang and Rhys; detailed calculations have not as yet appeared. Since, however, an increase in the transition probability of at least several dozen powers of ten would be necessary to account for the experimental results in a typical case, it is believed that other mechanisms might account for the radiationless transition.

One alternative mechanism has been suggested by Seitz[7] for the case of excitons. If the lattice relaxation is sufficiently great, the minimum of the excited state curve will fall outside of the ground-state curve. (See Fig. 1.) Efficient luminescence would then be impossible because the transition to the ground-state curve would occur either as the system approached its new equilibrium after absorption or by a thermal activation after reaching equilibrium at point *A*. Using experimental data on the *F*-band absorption, Russell has computed the configuration coordinate curves for the *F*-center in KCl, as shown in Fig. 2(a).[8] These curves indicate that Seitz's mechanism probably does not apply in the case of the *F*-center.

In the past, treatments of transitions from the excited electronic state have all been implicitly based on the assumption that the system is in thermal equilibrium (except for the quantum of electronic excitation energy). Consequently, the population of the highly excited vibrational states is assumed negligible in the computation of transition probabilities from the excited elec-

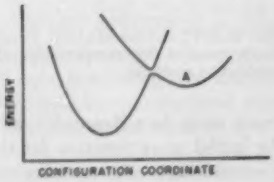


FIG. 1. An arrangement of configuration coordinate curves proposed by Seitz which would lead to radiationless transitions.

* Research supported in part by the U. S. Air Force through the Office of Scientific Research of the Air Research and Development Command.

[1] K. Huang and A. Rhys, Proc Roy. Soc. (London) **A204**, 406 (1950).

[2] S. I. Pekar, J. Exptl. Theoret. Phys. (U.S.S.R.) **20**, 510 (1950); **22**, 641 (1952).

[3] C. C. Klick, Phys. Rev. **94**, 1541 (1954).

[4] Botden, Van Doorn, and Haven, Philips Research Repts. **9**, 469 (1954).

[5] J. J. Markham, Phys. Rev. **94**, 1395 (1954).

[6] H. J. G. Meyer, Physica **20**, 181 (1954).

[7] F. Seitz, Trans. Faraday Soc. **35**, 74 (1939).

[8] G. A. Russell (to be published).

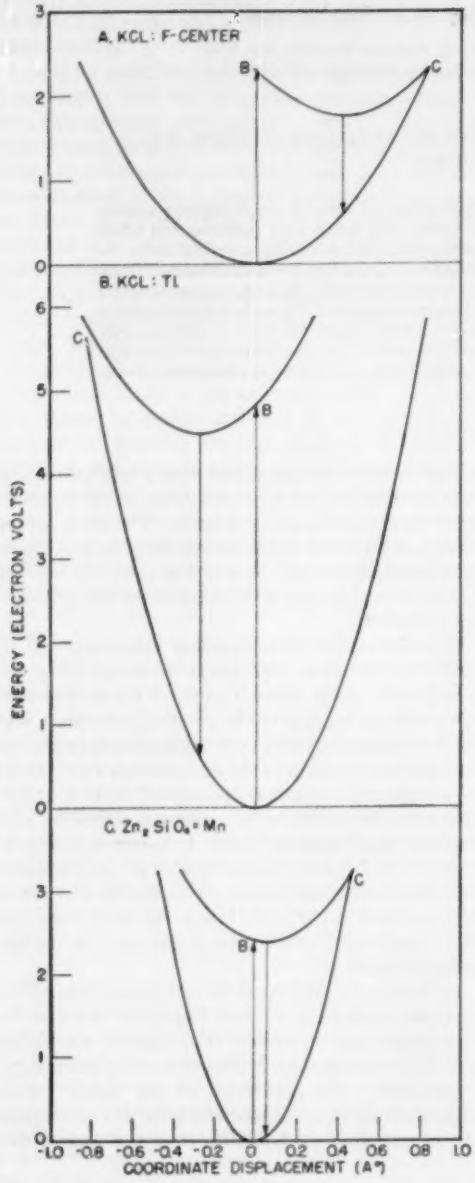


FIG. 2. Configuration coordinate curves for the F-center in KCl, KCl:Tl, and Zn_2SiO_4:Mn. In each case the coordinate is the displacement of the nearest neighbor ions from their ground-state equilibrium positions.

tronic state. In a Huang-Rhys-Lax[9] type of calculation, the initial wave function for the system in an excited electronic state thus contains very few or no phonons at low temperature. When pictured on a configurational coordinate diagram, the system is assumed to be in the vicinity of A on Fig. 3.

In the present note it is pointed out that the system, initially excited to the vicinity of B by the absorption of a quantum of light $h\omega$, may never reach the equilibrium position A because an alternative mechanism for nonradiative transitions may occur before the system comes to equilibrium, and that this alternative mechanism may be the basis of a criterion for luminescence in simple systems.

When the system is first excited by the absorption of $h\omega$, the system is in the Nth vibrational state, characteristic of the energy E_B-E_A. This energy will be radiated away in the form of phonons in a time long compared with 10^{-13} sec, if N is large. During this time, the configurational coordinates of the system will undergo violent oscillations, characteristic of the high local temperature. As the center "cools," the state of the system passes through the point with energy E_C and configurational coordinates X_C. At this point the energies of ground and excited electronic states are essentially the same, so that no activation energy is required for a redistribution of the electronic wave function, that is, for a transition to the ground electronic state. Thus before the system comes to its equilibrium position from B, it has the opportunity to change its electronic state at C; if the system once reaches A, however, it must await the unlikely occurrence of a statistical fluctuation in the local temperature to that characteristic of E_C, or must simultaneously emit a large number of phonons (to reach the point D), in order to change its electronic state by nonradiative transitions.

On the other hand, if the energy configuration is such that the intersection C occurs at an energy higher than that of B, as would occur if the two minima in Fig. 3 were laterally displaced only a short distance (small relaxation energy E_B-E_A), then during cooling the intersection point C would not be reached, and the system would reach A, where at not too high a temperature, it would stay long enough to luminesce. Thus a sufficient condition for luminescence at low temperatures is that E_C be greater than E_B; if the crossover at C occurs with high probability, the condition $E_C>E_B$ is also a necessary condition. We propose that this is sometimes the case.

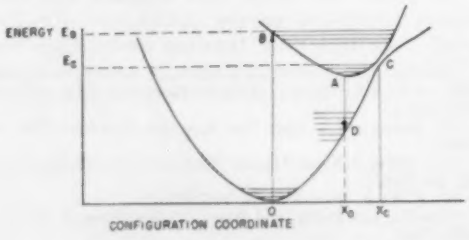


FIG. 3. Configuration coordinate curves leading to radiationless transitions.

[9] M. Lax, J. Chem. Phys. **20**, 1752 (1952).

In Fig. 2 are given the three systems for which configuration coordinate curves have been determined.[10] In the case of KCl:Tl[11] and Zn_2SiO_4:Mn[12] the centers are luminescent with high efficiency and the intersection of the curves (C) is well above the region to which optical transitions are made (B). In the case of the F-center, work of Russell [see Fig. (2a)], indicates that (C) is at least no higher than (B) in agreement with the low efficiency of luminescence in the F-center.

The criterion proposed above is not only in agreement with the experimental facts for the three known systems, but from it may also be derived a relationship between absorption and emission wavelengths which may be checked in a large number of systems for which the configurational coordinate curves have not been determined. On the basis of the criterion proposed here it is possible to determine the smallest value of the emitted energy in relation to the absorbed energy. The smallest emission energy occurs when C is at the same energy as B, since for C lower than B it is postulated that no emission occurs. Under these conditions the position of the minimum of the excited state curve on the coordinate axis falls halfway between the minimum of the ground-state curve and the intersection point. Then if the configurational coordinate curves are parabolic, and if $E_B = E_C$, the energy of the peak of the absorption band is

$$E_{abs} = \tfrac{1}{2} k_g (2X_0)^2, \tag{1}$$

where k_g is the force constant in the ground state and X_0 the lateral displacement of the two minima in Fig. 3. The energy of emission is given by

$$E_{emis} = E_B - \tfrac{1}{2} k_g X_0^2 - \tfrac{1}{2} k_e X_0^2, \tag{2}$$

where k_e is the force constant in the excited state. The values of k_g and k_e are not known in most cases but it is generally agreed that $k_e \leqslant k_g$. The minimum value of E_{emis} would occur if $k_e = k_g$. Thus we obtain

$$E_{emis} \geqslant E_{abs}/2, \tag{3}$$

or, in terms of wavelength,

$$\lambda_{emis}/\lambda_{abs} \leqslant 2. \tag{4}$$

[10] Configuration coordinate curves for $CaWO_4$ have been given by H. F. Hameka and C. C. Vlam [Physica **19**, 943 (1953)] and by C. C. Vlam [Brit. J. Appl. Phys. **5**, 443 (1954)] using a vibrational frequency obtained from Raman spectra in the determination of the curves. The resulting curves predict an absorption band at 3700 A which is not observed experimentally. For this reason the $CaWO_4$ data have not been included here although they do not disagree with the present criterion.

[11] F. E. Williams, J. Phys. Chem. **57**, 780 (1953).

[12] C. C. Klick and J. H. Schulman, J. Opt. Soc. Am. **42**, 910 (1952).

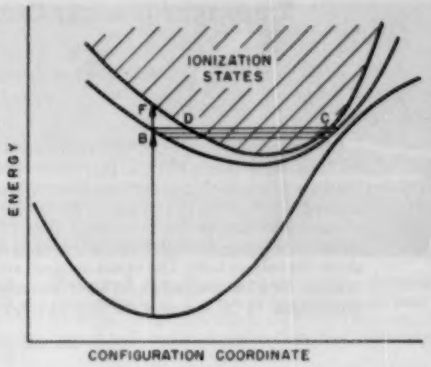


FIG. 4. An arrangement of configuration coordinate curves which might lead to photoconductivity without direct transitions to the conduction band. The upper curve represents the lowest energy ionization states of the center.

This inequality is strengthened further if one takes into account the deviations from the parabolic shapes, i.e., the increased curvatures, which are to be expected at large distances from the minima.

This upper limit for the wavelength ratio applies to a simple single center in which both the absorption and emission acts involve the same levels. A literature search reveals no such case where the relationship of Eq. (4) is violated. The general validity of the relationship of Eq. (4) gives added support to the mechanism of radiationless transitions from which it was derived.

There appears, then, to be a simple explanation for the low efficiency of luminescence of the F-center on the basis of configuration coordinate curves. The radiationless transition mechanism described here may be of general applicability for the case where absorption and emission transitions are localized at a center.

Another possible consequence of this mechanism might be the ionization of the center, that is, the removal of the electron, by absorption of a quantum of light appreciably less energetic than that predicted to be necessary by the Franck-Condon principle. If the ionization states dip below the point B at their minimum energy, a transition might also occur to them while the center is cooling. This possibility is indicated in Fig. 4, where the crossover might occur, energy-wise, at any configuration between X_C and X_D during the cooling process, although a direct Franck-Condon transition from the ground electronic state would require a photon of energy E_F, which may be appreciably larger than E_B. Thus photoconductivity might arise even at the lowest temperatures upon irradiating into discrete absorption lines as well as into bands associated with transitions directly into the conduction band.

Trapping of Minority Carriers in Silicon. II. *n*-Type Silicon

J. R. Haynes and J. A. Hornbeck
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received July 14, 1955)

Experimental evidence of temporary trapping of holes in the volume of *n*-type silicon in at least two different traps is presented. The trap parameters of cross section for capture and energy level above the valence band are determined by measuring the unfilled trap density, the hole density, the rate at which holes are trapped, and the rate at which holes are released from traps. Measurement of the temperature dependence of these quantities for the deeper traps gives, in addition, the temperature variation of the cross section for capture, the energy level of the traps below the conduction band and a second determination of the trap level above the valence band. The values obtained are consistent with accepted values of the energy gap. There is evidence for a loss mechanism for holes (recombination) from the deeper traps at a rate which is found to be proportional to the square of the electron concentration.

INTRODUCTION

In Part I of this paper,[1] we discussed the trapping of electrons in *p*-type silicon. It is the purpose of this section to present the results of our studies of the trapping of excess holes in *n*-type silicon.

Like the electron traps, hole traps must be crystal imperfections or impurities which can immobilize minority carriers and hold them in an unrecombined state for relatively long periods of time.[2] Also like the electron traps, hole traps evidence themselves by an increased conductivity that is associated with added majority carriers which exactly neutralize the space charge of the trapped minority carriers. Unlike the traps in *p*-type silicon, however, traps in *n*-type silicon produce a decay in photoconductivity which is very nearly a simple exponential function of time, showing that multiple trapping is not an important mechanism in this case. That is, an excess hole in the valence band is much more likely to be caught in a recombination center—and disappear—than to be caught in a trap. The absence of multiple trapping means that *n*-type silicon is a simpler trapping system to study than is *p*-type. Mainly for this reason we have made more extensive measurements of the temperature dependence of the trap parameters using *n*-type specimens.

n-type silicon differs from *p*-type also in the number density of traps that are found. In single crystals of *n*-type silicon which we have studied, trap densities have varied from less than $2\times10^{8}/cm^{3}$, i.e., an undetectable density, to approximately $10^{15}/cm^{3}$. Unlike *p*-type silicon, in *n*-type no correlation is found between resistivity and trap density. There is, however, the same correlation as in *p*-type between trap density and rotation of the crystal relative to the crucible when the crystal is grown. It should be emphasized that the traps discussed here occur "naturally" in the silicon as grown into single crystals at these Laboratories.

The trap densities found in *n*-type crystals can be so large that the resistivity in the dark of the crystal is affected appreciably by the trap's presence. In other words, the β traps behave, as we shall see, as an added set of donor (or acceptor) levels.

Measurements of the resistivity variation with temperature caused by "Fermi level filling" of the traps, together with data on the temperature variations of the capture cross section and mean-time-in-a-trap, fit together into a self-consistent picture of trap energy depth, and further, provide experimental proof that the principle of detailed balance applies to the traps.

KINDS OF TRAPS IN THE VOLUME OF *n*-TYPE SILICON

We have been able to isolate with certainty two kinds of traps in the volume of *n*-type silicon, and there is considerable evidence for the existence of at least two more trapping levels. The two prominent sets of traps occur in nearly all of our samples of *n*-type silicon grown by the crystal pulling technique of Teal and Little.[3] We shall call these centers α traps and β traps.

The α traps in 25 ohm cm material at room temperature are characterized by a time constant of the decay of photoconductivity of 45 milliseconds while the β traps have a decay constant of 300 seconds.

α Traps: Experimental Techniques and Methods of Measurement

Density and Time Constants

The α traps were studied with the aid of the chopped light and dc oscilloscope equipment used to investigate the shallow traps in *p*-type silicon described in Part I. The samples used were cut from single crystals into rectangular rods having dimensions $0.2\times0.2\times2$ cm and subsequently plated on the ends with rhodium followed by copper. A typical oscillogram obtained using an *n*-type silicon sample (42 A) with this apparatus is shown in Fig. 1. This is a reproduction of a photograph of the oscilloscope tube showing a trace of photoconductivity as a function of time. The sample is illuminated from A to C and the remainder of the time is in the dark. The conductivity rise from A to B is caused by the

[1] J. A. Hornbeck and J. R. Haynes, Phys. Rev. **97**, 311 (1955).
[2] These traps should not be confused with ordinary recombination or "deathnium" centers wherein recombination or re-emission occurs in a time less than the free time in the conduction stream.
[3] G. K. Teal and J. B. Little, Phys. Rev. **78**, 647 (1950).

excess electrons and holes produced by the light in the conduction and valence bands. These added carriers reach an equilibrium concentration in a time which is the order of the lifetime of holes in this material (40μ sec), which is short compared to the resolving time of the apparatus. The finite slope shown by the oscillogram is associated with the rise time of the light pulse.

A further, nearly linear increase in photoconductivity (from B to C) follows the initial rise. This photoconductivity change is evidence of the trapping of holes in α traps. At C, the light is extinguished and the conductivity suddenly drops to D. This drop, which is equal in magnitude to AB, reflects the decrease in carrier density caused by recombination of excess holes and electrons in the valence and conduction bands, i.e., a normal recombination process. A negligible number of traps have emptied in this length of time, however, so that the conductivity at D is higher than that at A by very nearly the difference between C and B. The line GHI was obtained approximately 0.1 sec after the illumination was cut off. It represents the conductivity of the specimen after practically all of the electron-hole pairs in the conduction stream have recombined and essentially all of the α traps have emptied. The sample still has an increased photoconductivity caused by some filling of the β traps, but no measurable decay of the β traps occurs on this time scale. Hence, we use the line GHI as a base line from which to measure the photoconductive decay time associated with the emptying of the α traps. This decay is very closely exponential with a time constant of 45 milliseconds.

If the intensity of illumination is doubled, the conductivity changes AB, CD, and BC increase nearly proportionally. With increasing illumination, however, a point is reached where, although AB and CD increase, BC does not. Under these conditions the trace BC is no longer linear but rises exponentially to a value which cannot be exceeded with further increase in illumination. Evidently this saturation is associated with the filling of all of the normally empty α trapping centers. As in Part I, we compute the density of α traps from the saturated value of the conductivity change DH, since for each trapped hole there is one added mobile majority carrier: $N_\alpha=(\Delta\sigma_{DH})_{\max}q\mu_-$, where q is the electronic charge and μ_- is the drift mobility of electrons. For specimen 42 A, which we are using as an example, we find $N_\alpha=1.4\times10^{12}$ cm^{-3}.

Capture Cross Section

The cross section for capture of holes by α traps may be calculated from data such as those shown in Fig. 1. The calculation makes use of an approximate formula which is derived in the Appendix from the general trapping equations (1) of Part I. The essential feature of the approximation is the interpretation already given of Fig. 1, *viz.*, that after a weak light is turned on, the excess hole density p rises quickly to very near its steady state value p_0 (the rise AB), and that subsequently

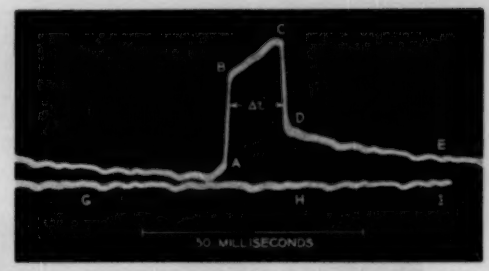


FIG. 1. Photograph of oscilloscope trace showing photoconductivity as a function of time. The sample was illuminated from A to C and the remainder of the time was in the dark. The conductivity difference from H to D was produced as a result of holes caught in α traps.

trapping in α traps begins to appear. The initial trapping rate dp_1/dt, given by the initial slope of BC, is closely equal to $(N_\alpha Sv)\,p_0$, where S is the trapping cross section, v is thermal velocity, $(N_\alpha Sv)$ is the trapping rate, and p_0 is the reservoir of holes from which the trapping occurs. Obviously p_0 is obtained from the measured conductivity change σ_{AB} by the relation $p_0=\sigma_{AB}/q(\mu_++\mu_-)$, and $dp_1/dt\simeq(\Delta\sigma/\Delta t)_{BC}(1/q\mu_-)$ where by $(\Delta\sigma/\Delta t)_{BC}$ we mean the initial slope of BC. With weak sources of illumination the change BC can be made linear so that there is little problem in determining this slope. These amplitude and slope measurements together with a knowledge of the mobilities and the trap density N_α yield S. For specimen 42 A we find $S_\alpha=4\times10^{-18}$ cm^2.

This value[4] is five orders of magnitude less than the capture cross section for traps in p-type silicon, which accounts for the absence of multiple trapping in the present case. More quantitatively, this may be illustrated as follows. The measured rate of recombination from the valence band of 2.5×10^4 sec^{-1} is about 400 times greater than the trapping rate $(NSv)=60$ sec^{-1} from the valence band. Thus a hole trapped in an α trap in this specimen has only one chance in 400 of being trapped again, even when all the traps are empty. Thus the time constant associated with the photoconductive decay DE is also the time constant τ associated with the emptying of the α traps. The emptying process could be either thermal ejection back to the valence bond band with subsequent normal recombination, or direct recombination in the trap, or a combination of these. If recombination in the trap is involved, we might expect the time constant of photoconductive decay to be a function of the majority carrier density.[5] The time constant, however, is found to be independent of majority carrier concentration, at least in this conductivity range. We, therefore, believe that the measured time

[4] This value was erroneously published as 40×10^{-16} cm^2 in J. R. Haynes and J. A. Hornbeck, Phys. Rev. 94, 1438(A) (1954).

[5] As shown in Part I, the time constant involved in recombination of trapped minority carriers in deep traps for p-type silicon varies as the square of the majority carrier density. This same law will be shown for the β traps in n-type silicon.

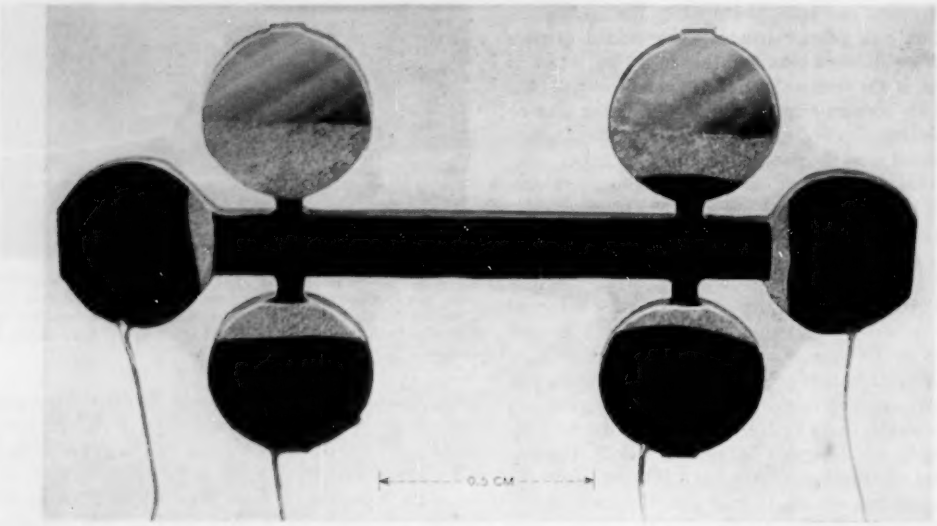


FIG. 2. Photograph of sample 376D-2. The etched surface of the silicon is visible in the central portion of the sample. The upper middle circular sections show the plated areas (top) and the sand blasted regions (bottom). The other four circular sections (with wires attached) are coated with black enamel.

constant τ is the time constant τ_g associated with the rate of generation to the valence band.

Energy Level

The energy level of the α traps with respect to the valence band was calculated from the experimentally determined values of S and τ (assumed equal to τ_g) using the detailed-balance formula, Eq. (1), which is discussed later. The result is that the α traps are located $\sim$0.45 ev from the valence band or $\sim$0.63 ev from the conduction band if the energy gap is 1.08 ev wide. For a resistivity of 25 ohm cm, the Fermi level lies 0.30 ev below the conduction band at room temperature, so that the α trapping centers are normally empty of holes.

β Traps: Experimental Techniques and Methods of Measurement

Experimental Techniques

Because of the long time constant associated with the decay of photoconductivity produced by β traps, it is essential in the study of these centers that the silicon samples be allowed to reach equilibrium in a close approach to absolute darkness.[6] An additional requirement, due to the high-temperature coefficient of resistance of silicon, is that the temperature of the sample remain essentially constant during the decay in photoconductivity. Both of these requirements were met by enclosing the samples in concentric, light tight boxes, with the space between the boxes thermally insulated, and finally placing the assembly in a Cenco-DeKhotinsky constant temperature furnace. Temperatures above room temperature were obtained by heating the furnace. Temperatures below room temperature were achieved by placing predetermined amounts of dry ice in the "furnace."

To produce photoconductivity, the samples were illuminated by the filtered light from a small flash light lamp placed in the inner box. The filter used was an optically polished silicon plate 0.07 cm thick. Light passing through this filter is deeply penetrating and produces electron-hole pairs throughout the volume of the sample. Variation of the intensity of the filtered light by known amounts was achieved by a calibrated resistance in series with the lamp.[7]

The silicon specimens were single crystals nearly all of which were cut into rectangular rods having similar dimensions to sample 42A (described earlier) and plated in the same way on either end. These two-electrode samples were placed in one arm of a Wheatstone Bridge circuit with a dc amplifier and pen recorder substituted for the usual galvanometer. The bridge was balanced with a sample in the dark. On closing a key, the light from the flash light bulb irradiated the sample. A record of the resultant bridge unbalance and subsequent slow decay of photoconductivity was produced by the pen recorder. Time constants longer than 10 seconds were recorded with a Leeds and Northrup stabilized dc amplifier and an Esterline-Angus recorder. Shorter

[6] The traps integrate the effect of all of the light which illuminates the sample during approximately a time constant. This may be a matter of several hours.

[7] The light was calibrated with a silicon p-n junction in terms of relative concentration of electron-hole pairs produced.

time constants were recorded with a Sanborn amplifier and recorder.[8]

With two-electrode samples there always exists the possibility that spurious resistance is produced at the electrodes.[9] In order to test the extent of the errors introduced in this way, a sample (376D-2) was cut as shown in Fig. 2. The circular sections were sand blasted and plated over half their area, first with rhodium and then with copper as shown on the upper two middle electrodes. Copper wires were then soldered to the two end and lower middle electrodes, which were finally given a thick coating of black glyptol.[10]

This four-terminal specimen was used in a circuit shown in Fig. 3. Current flows through the end of electrodes of the sample producing a potential difference between the two middle electrodes. This potential difference is balanced by an equal and opposite voltage produced by current flow through a resistance. The dc amplifier, with an input impedance of 20 megohms, and recorder give a trace of the voltage unbalance produced by a photoconductivity as a function of time.

The measurements obtained with this four-terminal sample were compared with those using a two-terminal sample (376D-1) which was cut from an adjacent region of the same silicon crystal. This comparison shows that the two samples give the same conductivity values except at the lowest temperature used ($ca-30°C$) where only minor differences occur. Because of this result, most of the measurements taken at room temperature were made with two-electrode samples while all measurements involving lower temperatures were made with the four-electrode specimen.

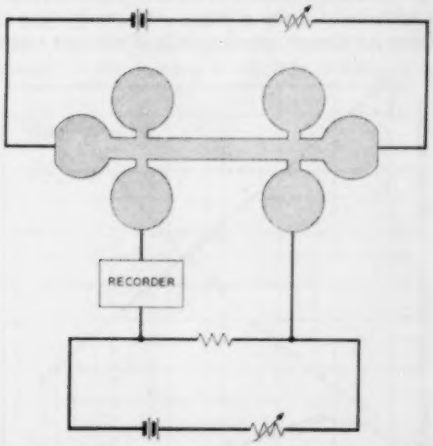


FIG. 3. Schematic diagram of circuit used with sample 376D-2. Current flows through the end electrodes and the resultant potential difference between the middle electrodes is balanced by the lower circuit. A permanent record of conductivity changes is obtained by the pen recorder.

[8] The time constant of the L&N Amplifier and E-A Recorder is ~0.5 sec. The time constant of the Sanborn is less than 0.01 sec.

[9] With silicon at liquid nitrogen temperature, such effects are known to be large.

[10] The purpose of the glyptol, which extends more then a millimeter beyond the plating (approximately 10 diffusion lengths), is to prevent the minority carriers produced by light from diffusing to the electrodes and possibly producing spurious potentials.

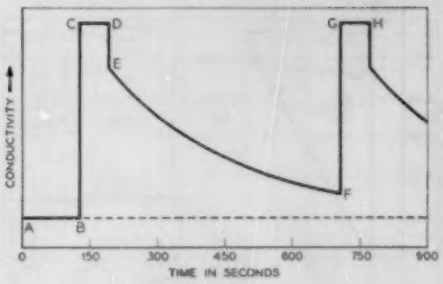


FIG. 4. Schematic of pen recorder chart of sample conductivity as a function of time obtained with relatively strong light. The sample was illuminated from B to D and from F to H, and the remainder of the time was in the dark. Analysis of such charts gives the mean time which holes remain in β traps and their concentration.

Measurement of Density and Time Constants

Both the density of normally unfilled β traps and the time constant of the decay in photoconductivity can be obtained from a pen recorder trace of photoconductivity as a function of time. A typical record made at room temperature is shown in Fig. 4. The trace from A to B is obtained with a sample in the dark and in equilibrium. At B, the sample is illuminated and in consequence its conductivity rises rapidly to a new value CD. At D, the light is extinguished and the conductivity drops to E as the holes in the valence band recombine and the α traps empty. This process is virtually complete in a few tenths of a second, during which time the concentration of holes in β traps remains essentially unchanged. From E to F, the β traps empty and the conductivity drops as the space charge neutralizing electrons in the conduction band disappear. At F, the sample is again illuminated with the same light intensity. The new conductivity represented by GH gives a check on the zero drift of the apparatus, a necessary precaution when long time constants (up to 10 000 sec) are measured. With no apparatus drift, the decay from E to F represents a true decay of the photoconductivity, and the conductivity difference BE results from the filling of some or all, if the light intensity is high enough, of the normally empty β traps, the density of which is given by this conductivity difference divided by $q\mu_n$. Normally only a very weak light is required to fill all of the β traps. In order to assure that this condition was fulfilled, a second recording was made with the light intensity increased by an order of magnitude. In this way, we measured for the β traps both N_β and the decay time τ as a function of temperature.

Measurement of Capture Cross Section

We shall next discuss the experimental measurement of the capture cross section of the β traps. As in the case

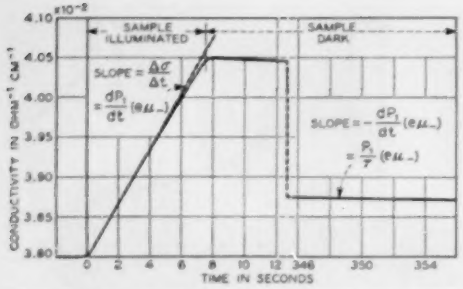


FIG. 5. Sample conductivity as a function of time obtained with relatively weak light and used in the determination of the capture cross section of β traps.

of the α traps, the method involves suddenly illuminating the specimen with weak, penetrating light and measuring the initial rate of trapping. Figure 5 is a typical recorder trace showing the sample conductivity as a function of time. The important difference between Fig. 5 and Fig. 1, which pertains to α traps, is that at $t=0$ no abrupt change in conductivity is observed corresponding to $p_0 \simeq l/\tau$, the rise AB in Fig. 1. This is explained by the fact that in the β trap measurement the light intensity is so weak that p_0 is negligible ($\sim 10^8$ or 10^9 cm^{-3}) compared to the number of holes in β traps after a time corresponding to the time resolution of the apparatus, a few tenths of a second. In this time filling of the α traps reaches a steady state, but a negligible number of holes are caught there.

Lacking the amplitude change corresponding to p_0, we must measure l and τ independently. The value of l was determined by using an n-type sample having the same dimensions and surface treatment as sample 376D-2 (the specimen studied most extensively), but which contained no detectable density ($<10^8$ cm^{-3}) of either α or β traps. When this specimen was illuminated with the weak light source, its photoconductivity is caused entirely by additional electron-hole pairs in the conduction and valence band, respectively, and hence $l=p/\tau_r$, where p comes from a measurement of the photoconductivity and τ_r is the measured life time in the sample.

The value of l obtained in this way was checked by substituting a p-type sample, having the same dimensions and surface treatment, which contained many deep traps. In this sample, nearly all the electrons produced by the light are trapped before they can recombine, i.e., the sample integrates the light falling on it for weak incident illumination and times short compared to its decay time.

The value of τ for specimen 376D-2 was obtained in a straightforward way as described in Part I. With these procedures and the formulas described[11] in Appendix I, the capture cross section of the β traps was measured as a function of the temperature.

[11] Corrections to the simple formula used in the case of the α traps are discussed in the Appendix. There the right hand, or long time, portion of Fig. 5 is discussed.

EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present the results of the measurements as a function of temperature on the β traps. We shall show that the measurements of the β trap density, N_β, determine the energy level of the traps with respect to the conduction band, that measurements of the capture cross section together with measurements of the decay time, τ, when interpreted properly determine the trap energy level with respect to the valence band, and that these results are consistent with the principle of detailed balancing. From this interpretation, we also obtain information on the rate of recombination in the β traps.

Capture Cross Section

Using the procedure described in the last section, we have determined the capture cross section S of the β traps as a function of temperature T. The results are shown in Fig. 6, which is a plot of ln S as a function of T^{-1}. The circles represent values of S obtained by using the n-type silicon sample (containing no measurable trap density[12]) to evaluate l, whereas the squares represent values obtained by using the p-type sample for the same purpose. Within experimental error the data fall on a straight line represented by $S=S_0 \exp(-2100/T)$ with $S_0=8\times10^{-15}$ cm^2. The activation temperature 2100°K corresponds to an energy of 0.18 ev.

At room temperature, $S=8\times10^{-18}$ cm^2, a value twice that obtained for the α traps. The small value of S explains the almost complete lack of multiple trapping

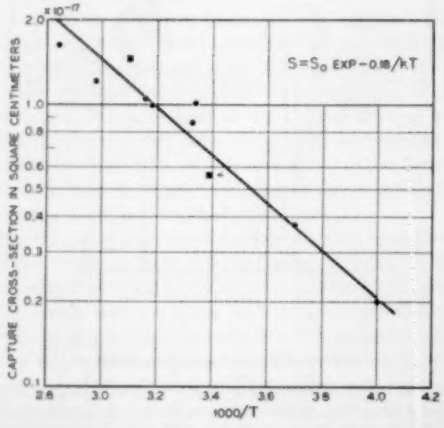


FIG. 6. Capture cross section of β traps as a function of temperature. The solid circles were obtained using a sample containing no detectable traps for the determination of l. The squares were obtained using a p-type sample with a high trap density for this purpose.

[12] Although we found no measurable density of volume traps in this specimen, we did find that after heating the specimen in an oven in the course of the experiments, small trapping effects showed up that were associated with the surface. If unnoticed, these effects would have affected our experimental results.

in β traps. In specimen 376D-2 at 35°C, the rate of recombination from the valence band is ten times the trapping rate[13] ($NS\nu$). In the neighborhood of room temperature in this specimen, a trapped hole has about one chance in ten of being retrapped when all the traps are empty; hence the final decay rate is about 10% longer than the initial decay rate when most of the traps are filled, and there is a small but detectable amount of multiple trapping in β traps.

Energy Level with Respect to the Conduction Band

The energy level of β traps with respect to the conduction band can be obtained by a comparison of the trap level and the Fermi level. This is most easily done with a specimen (such as 376D-2) which has a large trap density and a resistivity which places the Fermi level at room temperature close to the energy level of the β traps. Under these circumstances the sample shows an anomalous behavior of conductivity as a function of temperature. Such data are shown in Fig. 7, which is a plot of sample conductivity as a function of temperature for two conditions, (a) with the sample in equilibrium in the dark, and (b) with all of the β traps filled by illumination. The conductivity obtained with this sample in the dark (curve a) has two inflection points and is quite different from a sample containing relatively few β traps. These "normal" samples show dark conductivity in this temperature range which decreases contiuously as the temperature is increased, as does curve *b*. If essentially all of the impurity atoms are ionized, as is the case for these samples, the decrease in conductivity is caused by the decrease in electron mobility with increasing temperature. The data defining this upper curve, however, are values also attained with sample 376D-2, but with all of the β traps filled with holes produced by illumination, as described previously. From Fig. 7, the concentration, N, of normally empty β traps as a function of temperature shown is therefore given by $(\sigma_b-\sigma_a)/q\mu_-$.

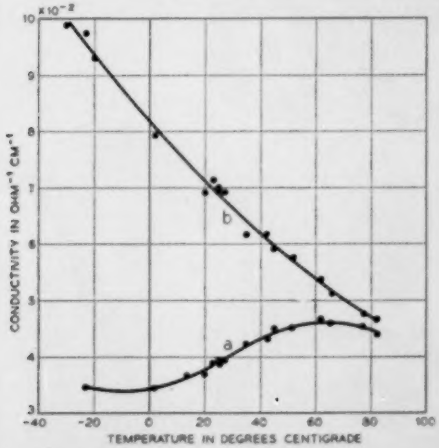


FIG. 7. Conductivity as a function of temperature for sample 376D-2 for two conditions; (a) with the sample in equilibrium in the dark and (b) with the β traps filled with photo-holes.

[13] At lower temperature this latter rate drops because of the decrease in S, and at higher temperature it drops because of the decrease in N.

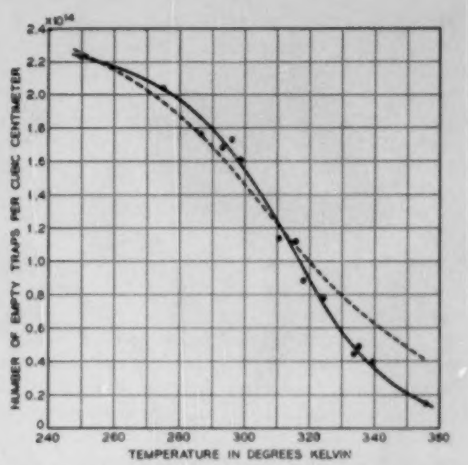


FIG. 8. Density of normally empty β traps as a function of temperature. The solid line is drawn through the experimental data. The dashed line is the expected variation of trap density with Fermi level change. See text for assumptions.

Values of N obtained in this way are shown as circles in Fig. 8 in which we have plotted the normally unfilled trap density as a function of absolute temperature.[14] The trap density may be seen to approach a maximum value of $\sim 2.4\times10^{14}$ cm^{-3} at low temperature and to drop rapidly in the region of room temperature so that at slightly elevated temperatures (80°C) less than one tenth of the trapping centers remain normally unfilled with holes. This change of trap concentration with temperature is, of course, associated with movement of the Fermi level.

The dashed line in Fig. 8 is an approximate calculation of the variation with temperature of the equilibrium density of empty traps obtained by making use of two well known relations: (a) the Fermi-Dirac distribution which relates the fraction of empty traps to the difference between the Fermi level and the trap energy level, and (b) the relation between the position of the Fermi level and the sample conductivity, which is known experimentally (curve *a* of Fig. 7). In this calculation it was assumed that the product $(g_0/g_1)(m_a/m)^{\frac{1}{2}}=1$, where (g_0/g_1) is the ratio of the

[14] The values of μ used were obtained by assuming $\mu=1200$ cm^2/volt second at 295°K and $\mu=$constant $T^{-2.4}$ as given by the experimental points of curve *b* of Fig. 7 assuming that the number of carriers is constant. This temperature dependence is also in accord with the exponent -2.6 found for pure lattice scattering by F. J. Morin and J. P. Maita [Phys. Rev. 96, 28 (1954)] with allowance for some impurity scattering.

statistical weight of the trap when it is empty to that when it is filled with a hole and (m_e/m) is the ratio of the effective mass of a conduction electron in n-type silicon to the mass of a free electron. It was assumed, further, that the temperature dependence of the trap energy level[15] was one half that of the total energy gap in silicon, i.e., that in electron volts $E_t(T)=E_t(0°\text{K})-2\times10^{-4}T$, where T is the absolute temperature, and that at 300°C, $E_t(T)=0.32$ ev below the conduction band.

The fit between the two curves, Fig. 8, is not as good as we expected, and we are unable to explain the discrepancy. The fit would be much worse if, instead of a single trapping level, we assumed a band of levels a few hundreds of an electron volt wide. More careful calculations[16] were made in which we assumed (a) the same temperature coefficient and mobility variation as above and (b) the "density of states" electron effective mass as computed from cyclotron resonance experiments,[17] and then looked for the best values of the unknown parameters for the two simple cases $(g_0/g_1)=\frac{1}{2}$ and $(g_0/g_1)=2$, corresponding to the traps being donor-like or acceptor-like, respectively. The donor case fitted better than the acceptor case, and yielded the following results: $E_t(0°\text{K})=0.35$ ev, N_0, the density of trapping centers $=2.6\times10^{14}$ cm^{-3}, and (N_D-N_A), the difference between the densities of "normal" donor and acceptor centers, $=1.1\times10^{14}$ cm^{-3}.

In any case, we believe the agreement between theory and experiment is close enough to warrant the conclusion that we are dealing with a single trap level and that its position relative to the conduction band is known with an uncertainty of only a few hundredths of an electron-volt.

Energy Level with Respect to the Valence Band

The energy level of the β traps with reference to the valence band can be determined by measurements as a function of temperature of the capture cross section and $1/\tau_g$, the rate of thermal release from the traps, and then invoking the principle of detailed balancing. Herring[18] has given a formulation of this principle which we shall employ, *viz.*,

$$1/\tau_g=16\pi m_h^*(kT)^2h^{-3}(g_0/g_1)S\exp[(E_v-E_t)/kT]. \quad (1)$$

Here the symbols are as previously defined with the additions that E_v is the energy at the top of the valence band and m_h^* is an effective mass for holes. Herring has pointed out to us that, strictly speaking, this formula is not applicable to our case because it does not take into account the facts that (1) the top of the valence band in silicon is degenerate, (2) the effective mass to be used at room temperature is different from that obtained from cyclotron resonance experiments because kT is the order of the energy associated with spin orbit splitting. He does not feel, however, that an appreciable error for our purposes will be incurred by ignoring these complications, setting $m_h^*(g_0/g_1)=1$, and employing experimental values of S and τ_g in Eq. (1).

This equation predicts that except for the T^2 effect a plot of $\ln 1/\tau_g$ *vs* $1/T$ should be a straight line. In Fig. 9, we have plotted as circles experimental values of $\ln 1/\tau$ *vs* $1/T$, where τ is the photoconductivity decay constant corrected for a small multiple trapping effect. The fact that these data do not lie on a straight line suggests that there is a loss mechanism from the trap in addition to thermal release to the valence band. A reasonable assumption here is that the additional loss mechanism is recombination of the trapped holes with conduction electrons. If $1/\tau_b$ is this recombination rate, then as in Part I,

$$1/\tau=1/\tau_g+1/\tau_b, \quad (2)$$

and we attempt to draw two straight lines the sum of which fits the experimental data. As shown in Fig. 9, this can be done, for the triangles which fall on a straight line were obtained by subtracting from the data the line having the lower slope. Either line might represent generation out of the trap. The line of lower slope, however, has an activation energy much too low to represent $1/\tau_g$, so we attribute it to $1/\tau_b$.

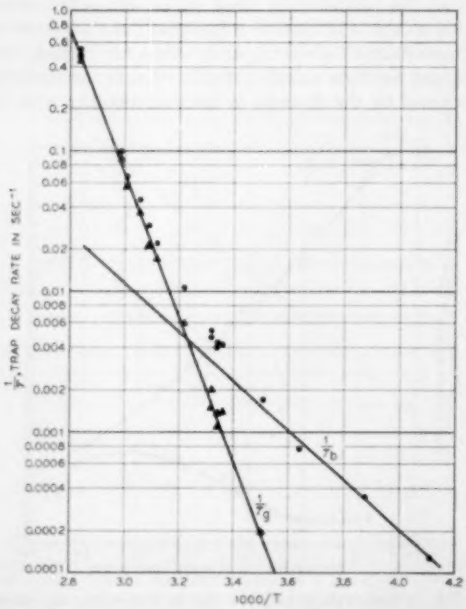


FIG. 9. Reciprocal of the time constant of decay of photoconductivity due to β traps as a function of reciprocal temperature. The data are shown by the open circles. The triangles are the result of subtracting the line of lower slope, labeled $1/\tau_b$, from the data.

[15] We are indebted to Dr. C. Herring for discussions of this point.
[16] We are indebted to Dr. M. C. Gray for carrying out these computations.
[17] R. N. Dexter and B. Lax, Phys. Rev. **96**, 223 (1954).
[18] C. Herring, *Fundamental Formulas of Physics* edited by D. H. Menzel (Prentice Hall Publications, Inc., New York, 1955), p. 630.

We may now use the detailed balance equation to determine (E_v-E_t) in two different ways from the data: (1) from slope of the line $1/\tau_g$ vs $1/T$ (with corrections for the T^2 factor) and (2) from the measured values of S and τ_g at a given temperature. From the slope of the line we obtain the absolute zero difference $(E_v-E_t)=0.79$ ev after taking into account the 0.18 ev activation energy associated with S. Correcting this value to 300°K by using the temperature coefficient discussed earlier (2×10^{-4}), we obtain $(E_v-E_t)_{300°K}=0.72$ ev. From the measured values of τ_g and S at 300°K we also obtain $(E_v-E_t)_{300°K}=0.72$ ev, which is somewhat better agreement than we expect. Thus we conclude that the detailed balance principle applies in a consistent fashion to the data on the τ_g process, which gives us confidence in the results and their interpretation.

It should be noted, further, that these results together with the conductivity data predict an energy gap in silicon at 300°K of $(0.72+0.32)=1.04$ ev. This value lies within experimental error of that of 1.1 ev recently obtained by Morin and Maita.[19]

Recombination

The line of lower slope in Fig. 9, which we have attributed to recombination, empirically is of the form $1/\tau_b=\nu\exp-E/kT$. From the data, $\nu=2.2\times10^8$ sec^{-1} and the activation energy $E=0.35$ ev. From the low value of ν it is evident that the process is an improbable one.

More information on the recombination process may be obtained from measurements of decay time on a variety of specimens at room temperature. Since $1/\tau_g$ is known, $1/\tau_b$ can be obtained from $1/\tau$ by use of Eq. (2). The results of these measurements are shown in Fig. 10, in which we have plotted the rate $1/\tau_b$ as a function of the concentration of electrons in the conduction band.

The solid circles were obtained using different specimens while the triangles were all obtained with a single sample. The triangle at the lower left was measured with the sample before heat treatment. The other triangles from left to right were obtained after heating the sample at 460°C for 1, 4 and 10 hours, respectively. This heat treatment increased the majority carrier concentration by more than an order of magnitude while leaving the concentration of β trapping centers substantially constant. The line drawn has a slope of two so that, within experimental error, the recombination rate is found to be proportional to the square of the majority carrier concentration. This is the same dependence found for the recombination rate of electrons in deep traps in p-type silicon (see Sec. I). It must, therefore, represent some general fundamental process.

It has been suggested that this process is the inverse of an Auger-like process in which two slow electrons in the conduction band interact with one of the electrons ending up in the trap, i.e., recombining with a hole, and the other is speeded up, conserving energy.

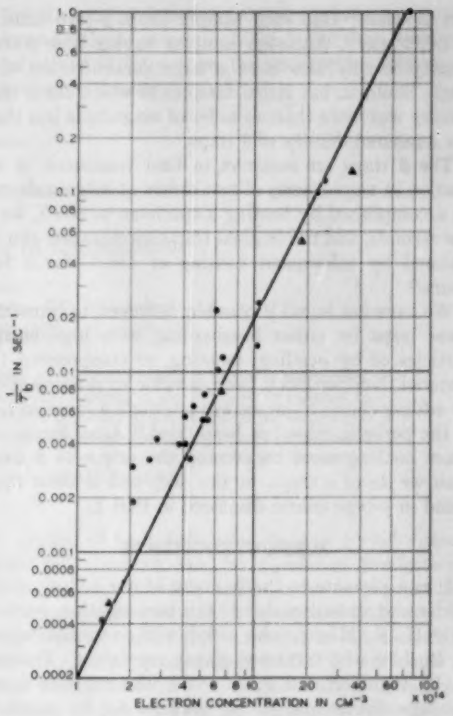


FIG. 10. Recombination rate for holes trapped in β traps as a function of majority carrier (electron) density.

The large spread in the points is to be expected since the carrier concentration was calculated from the average conductivity of the sample. Consideration shows that samples having a nonuniform conductivity will produce data which are shifted downward and to the right from those obtained using uniform specimens. A few data are not shown. These were obtained with samples having time constants nearly two orders of magnitude larger than expected. This large difference cannot be explained on the basis of nonuniform conductivity and it is, therefore, strong evidence for the presence of another deep trapping level (γ traps).

Factors Effecting Trap Density

Densities of α and β traps[20] usually lie in the range of 10^{13} to 10^{14} cm^{-3} for material grown by the method of Teal and Little (crystal rotated while grown) and are approximately two orders of magnitude less for crystals grown in the same way but without rotation. Samples having the lowest trap density are those produced by the floating zone crystal growing technique of Theuerer

[19] F. J. Morin and J. P. Maita, Phys. Rev. 96, 28 (1954).

[20] Most of the α-trap concentration data were obtained by R. G. Shulman.

and Whelan.[21] One such sample has a β-trap density $<2\times10^{9}$ cm^{-1}. Although samples having high β-trap density usually also have a large concentration of α traps, Shulman has found samples in which the α trap density was more than an order of magnitude less than our measured density of β traps.

The β traps are sensitive to heat treatment. A reduction in trap density of two orders of magnitude can be accomplished by heating a specimen to 800°C for a few seconds, and the original trap concentration can be restored by subsequent heating at 450°C for a few hours.[22]

We have not found it possible, however, to introduce these traps by either bombarding with high energy particles or by bending, twisting, or compressing the material. Neither has it been possible to produce them by adding chemical impurities although a large section of the periodic table has been tried.[23] Accordingly, we know nothing more concerning the origin of β traps than we do of α traps, or the deep and shallow traps found in p-type silicon discussed in Part I.

ACKNOWLEDGMENTS

It is a pleasure to thank many of our colleagues for advice and assistance during this investigation, particularly H. R. Moore, who aided with instrumentation, E. Buehler who furnished silicon crystals, C. Herring, G. H. Wannier, and P. A. Wolff who supplied many valubale discussions, R. G. Shulman for his measurements of α trap density, J. H. Wiley, who measured some of the time constants of β traps and W. C. Westphal who assisted us generally.

APPENDIX: APPROXIMATE FORMULAS FOR USE IN MEASURING CAPTURE CROSS SECTIONS

We are interested here in deriving simple, approximate formulae which may be used in determining the capture cross section S of the α and β traps. These derivations are based in part on the experimental conclusion that the rate of capture of holes in traps is small compared with their rate of capture in deathnium (normal recombination) centers.

We start with the equations of trapping kinetics for a single kind of center as given in the Appendix of Part I:

$$dp/dt=l-rp+gp_1-p(N-p_1)Sv, \tag{3}$$

$$dp_1/dt=-gp_1+p(N-p_1)Sv. \tag{4}$$

Here p = the excess density of holes in the valence band, p_1 = the excess density of holes in traps, $r\equiv1/\tau_r$ = rate of recombination of excess holes in the valence band, $g\equiv1/\tau_g$ = thermal rate of generation of holes from traps, S = cross section for capture of holes in traps, N = density of normally unfilled traps, v = thermal velocity, l = rate of generation of electron-hole pairs per cm^3 by external light, and NSv = rate of trapping of holes when all traps are empty.

Prior to $t=0$, let the specimen be in the dark and all traps empty (i.e., thermal equilibrium). For $t\geq0$ let a light shine on the specimen such that a constant number l of hole electron pairs is produced per cm^3 per sec. Near $t=0$, $p_1\ll N$, and if in addition $g\ll NSv$, then

$$\dot{p}\simeq l-(r+NSv)p, \tag{5}$$

$$\dot{p}_1\sim NSvp. \tag{6}$$

Integrating Eq. (5) gives

$$p\simeq[l/(r+NSv)][1-\exp(r+NSv)t]. \tag{7}$$

That is, in the order of time no longer than the recombination time $\tau_r=1/r$, the density of excess holes in the valence band increases to a value $p=l/(r+NSv)$. The final steady state value of p from Eqs. (1) is $p=l/r$. Thus if $r\gg NSv$, as is generally the case for n-type silicon, the density of holes in the valence band quickly rises essentially to its steady-state value and remains constant.

We have then from Eq. (6) that

$$\dot{p}_1\simeq NSvl/(r+NSv), \tag{8}$$

so that if we measure the initial slope of the change in conductivity with time milliseconds to seconds after the light is turned on, and before p_1 becomes appreciable compared to N, and also measure N, r, and l, then S can be gotten directly from Eq. (8):

$$S=r\dot{p}_1/Nv(l-\dot{p}_1).$$

If $r>NSv$ such that p does not change by more than say 10% after several lifetimes (τ_r), so that it may be known quite well, slightly more accurate values of S may be obtained by using Eqs. (3) and (4) rather than (5) and (6) for $\dot{p}_1$. For according to Eqs. (3) and (4), if p is constant, and we know that $l/(r+NSv)<p<l/r$,

$$\dot{p}_1\simeq NSvl/(r+NSv)-(g+pSv)p_1. \tag{9}$$

By keeping p_1 small so that it never exceeds a few percent of N, then the measurement of the initial slope $\dot{p}_1$ will give a reliable value of S unless g becomes large.[24] This can happen at high temperatures since g is increasing exponentially with T, e.g., $g\propto\exp(-8200/T)$. As long as NSv is still much smaller than r, so that p is essentially constant, we can correct for this g-effect by adding to the measured initial slope the negative of the slope of the decay curve (obtained after the light is turned off) for the same value of p_1, for this slope is

[21] H. C. Theuerer and J. M. Whelan, Trans. A.I.M.E. (to be published). These data have also been independently derived and described by Keck, Emis, Müller, and others: P. H. Keck, Physica **20**, 1059(1954); R. Emis, Z. Naturforsch. **9A**, 67 (1954); S. Müller, Z. Naturforsch. **9B**, 504 (1954).

[22] Hannay, Haynes, and Shulman, Phys. Rev. **96**, 833(A) (1955).

[23] This work was done in collaboration with N. B. Hannay and E. Buehler.

[24] Both correction terms in Eq. (5), viz., $(g+pSv)p_1$ can be computed from the experimental data and applied if one desires.

just gp_1. This is shown schematically in the right hand position of Fig. 5. It was necessary to make this correction only at the highest temperature (62°C) at which the cross section of the β traps was measured.

For the α traps, it is not necessary to measure l and τ independently in order to obtain S, for their ratio can be obtained directly from the oscillogram, Fig. 1, as described in the text.

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Precipitation of Copper in Germanium

R. A. LOGAN
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received June 27, 1955)

The density of dislocations is shown to have a marked effect on the rate of anneal of copper in germanium. At 500°C samples containing high dislocation density ($\sim 10^6$/cm^2) anneal in about 1 hour in contrast to material of low dislocation density ($\sim 10^4$/cm^2) which requires about 24 hours. When copper-doped germanium is cooled from a high temperature in regions of high dislocation density, significant precipitation occurs in a cooling cycle of only a few seconds. In this case, in order to prevent precipitation the sample must be quenched from the high temperature in a time of the order of 0.1 second.

INTRODUCTION

WHEN germanium is heated to high temperature without precaution to prevent chemical contamination, impurities diffuse into the crystal. Copper has an appreciable solubility and high diffusivity in germanium[1] and is generally the principal contaminant in most heating procedures. Copper is an acceptor in germanium with an energy level[2] 0.04 ev above the valence band and a higher energy level[3] near the center of the energy gap.

When germanium is heated at a given temperature in the presence of copper at equilibrium, the concentration of the solid solution is characteristic of the temperature only. Except for the temperature range near the melting point where retrograde solubility occurs,[4] the solubility increases with increasing temperature. Hence except for retrograde solubility effects, as soon as the temperature is lowered from this equilibrium temperature, the copper solution immediately becomes supersaturated; and the copper concentration, as inferred from resistivity measurements, will drop from the high-temperature value to that characteristic of the lower temperature. A study[1] of this annealing process with techniques using radioactive copper showed that the copper precipitated at various places in the crystal. The extent to which this copper precipitates depends mainly on the cooling cycle and the density of nucleation sites.

In the experiments reported here, it is shown that precipitation of copper is extremely rapid in regions of the crystal of high dislocation density ($\sim 10^6$ dislocations/cm^2). In such samples significant precipitation occurs in a few seconds, and in order to prevent precipitation the sample must be quenched from the high temperature in a time of the order of 0.1 second. In addition, complete annealing at 500°C will occur in about 1 hour in contrast to a time of about 24 hours in material of low dislocation density ($\sim 10^4$ dislocations/cm^2).

In these experiments, dislocations associated with plastic deformation are introduced into germanium by bending at high temperature. The copper solubility in these samples, determined from resistivity measurements, shows a marked dependence on the density of dislocations. In regions of high dislocation density ($\sim 10^6$/cm^2), the density of acceptors due to copper is much lower than in regions of low dislocation density ($\sim 10^4$/cm^2). This phenomenon gives rise to an apparent decrease in copper solubility (as inferred from resistivity measurements) in regions of high dislocation density.

EXPERIMENTAL PROCEDURE

In these experiments, rods of dimension 0.06 in.×0.1 in.×1.0 in. were cut from a single crystal of germanium which had a room temperature resistivity of about 0.8 ohm-cm, n-type. The rods were bent about the (111) crystal axis, without introducing chemical contamination, by a method devised by Pearson.[5] This method consists of (1) cleaning the rod,[6] (2) clamping it between molybdenum electrodes, (3) heating it by the joule heat developed by passing current through the rod, and (4) deforming the rod by bending at a high temperature (~800°C). The bent rods were then dipped in an

[1] Fuller, Struthers, Ditzenberger, and Wolfstirn, Phys. Rev. **93**, 1182 (1954).
[2] F. J. Morin and J. P. Maita, Phys. Rev. **90**, 337 (1953).
[3] Burton, Hull, Morin, and Severiens, J. Phys. Chem. **57**, 853 (1953).
[4] C. D. Thurmond and J. D. Struthers, J. Phys. Chem. **57**, 831 (1953).
[5] Pearson, Read, and Morin, Phys. Rev. **93**, 666 (1954).
[6] R. A. Logan, Phys. Rev. **91**, 757 (1953).

aqueous solution of $CuSO_4$ and heated for 5 minutes at 750°C to diffuse the copper into the crystal. At this temperature the diffusion constant of copper[1] is 2.8×10^{-6} cm²/sec, and in 5 minutes copper will thus diffuse a mean distance of about 40 mils.

The rods were then cleaned and heated in a furnace designed to heat and quench germanium.[7] In this and subsequent heatings precautions were taken[6] so that no further chemical contamination was introduced into the samples. The heating cycle employed was as follows: The sample was heated to 900°C in 1 minute and was maintained at this temperature for 1 minute. The sample was then cooled by one of two possible cycles. In the first cycle the power to the furnace was simply shut off, and the sample cooled to 500°C in about 15 seconds and to room temperature in about 2 minutes. In the other cycle the sample was quenched by dropping it out of the furnace into a bath of ethylene glycol. In the quench, it is estimated by considerations of heat conduction that the sample cooled to near room temperature in about 0.1 second.

Two rods of germanium were bent to radii of curvature of 2.5 cm and 20 cm, introducing dislocation densities of about 10^7/cm² and 10^6/cm², respectively.[8] The rods were n-type with $N_D - N_A \sim 2 \times 10^{15}$/cm³ so that acceptors associated with the added dislocations[5] would have a negligible effect on $N_D - N_A$. The rods were then cleaned[6] and heated to 900°C in the two-minute cycle as described above, followed by either a quench or a "slow" cool.

Resistivity measurements were then made along the length of the rod by the 2-point probe technique, using points separated by 0.010 in. Resistivity readings were recorded at steps of 0.0052 in., so that a resistivity profile of the rod was obtained. The resistivity type (n or p) was determined from the sign of the thermoelectric voltage (thermal probe technique).

RESULTS

The resistivity profile of the rod bent to $R = 2.5$ cm is shown in Fig. 1. Curve (a) shows the resistivity profile obtained after heating the rod to 900°C followed by a "slow" cool. Curve (b) shows the profile of the same rod obtained after quenching from 900°C.

It is evident that after quenching, the resistivity along the rod is nearly uniform, and the impurity concentration deduced from the resistivity measurements varies less than 15% from the mean value. However, in the case of the "slow" cool there is a marked resistivity variation along the length of the rod, and the associated impurity concentration varies by more than an order of magnitude. It will be shown in a subsequent paper that no acceptors associated with defects are introduced by the quench since the high dislocation density prevents the formation of these defects.

From the resistivity measurements the increase in acceptor density can be calculated.[9] In the region of maximum deformation more acceptors are found after quenching than after the "slow" cool. This effect is reversible in that heating to 900°C followed by either a quench or a "slow" cool will cause the acceptor density in the deformed region to cycle back and forth to the values characteristic of the quench and "slow" cool, respectively. This effect is presented in Fig. 2 where a plot is made of the added acceptor concentration measured after heat cycles which end with either a quench or "slow" cool. In this case, the room temperature resistivity was measured at the region of maximum deformation by the 2-point probe technique using points separated by 0.150 in.

Studies of low-temperature annealing at 500°C in plastically deformed germanium indicate that again the dislocation density has a marked influence on the rate of copper precipitation. Figure 3(a) shows the resistivity profile of the rod bent plastically to an $R = 20$ cm. This rod was heated 5 min at 750°C in the presence of copper and cooled to room temperature in about 1 minute. The rod had an initial resistivity of 0.8 ohm-cm, n-type, and after bending and introducing copper into the crystal, the crystal remained n-type in the region of the deformation. The ends of the rod which were held by the clamp type electrodes used to support the rod while bending remained cool and hence undeformed during the bending. In these regions of low dislocation density the rod converted to low resistivity, p-type.

This rod was annealed at 500°C in a nichrome furnace in an H_2 atmosphere. The heating was inter-

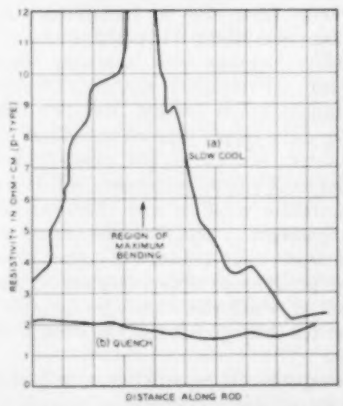


FIG. 1. Curve (a) is a plot of 2-point probe resistivity measurements along the copper doped bent rod ($R = 2.5$ cm) after heating to 900°C and "slow" cooling to 500°C in about 15 sec; to room temperature in about 2 min. Curve (b) shows 2-point probe resistivity measurements along the same rod after heating to 900°C and quenching in about 0.1 sec.

[7] This furnace is very similar to that described in 6.

[8] W. T. Read, Jr., *Dislocations in Crystals* (McGraw Hill Book Company, Inc., New York, 1953), p. 38.

[9] M. B. Prince, Phys. Rev. **92**, 681 (1953).

rupted after 5, 20, and 60 minutes, and room temperature resistivity profiles were obtained. These are shown in Figs. 3(b), 3(c), and 3(d), respectively. Two features of this anneal are noteworthy. In the region of maximum dislocation density ($\sim 10^6/cm^2$), essentially complete anneal has occurred in about 1 hour, whereas in the undeformed region, little annealing has occurred in this time. Secondly, although it is not understood, it is evident that in the first few minutes of anneal, more acceptors are present in the deformed region than at the beginning of the anneal.

DISCUSSION AND CONCLUSIONS

The results presented above show that the rate of precipitation of copper in germanium is increased markedly in the presence of large dislocation density. Although the actual mechanism of precipitation was not ascertained, it seems reasonable that copper atoms become attached to dislocations and act as centers of nucleation for

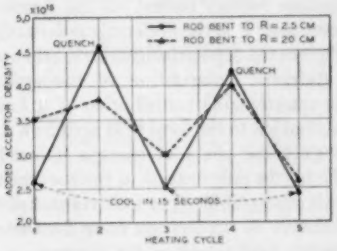


FIG. 2. The added acceptor concentration in the copper doped bent rods plotted after successive heat cycles as follows: 1, 3, and 5—heated to 900°C and cooled to 500°C in 15 sec; to room temperature in about 2 min. 2 and 4—heated to 900°C and quenched in about 0.1 sec.

precipitation of other copper atoms. The rate of anneal would also be markedly increased if copper atoms diffused along dislocations much faster than they diffuse in the pure crystal.

These results suggest an explanation to some early experiments on annealing of copper in germanium. It was observed[10] that specimens obtained from multicrystalline cast ingots annealed much faster than specimens of equal resistivity obtained from single crystals grown by conventional crystal-pulling techniques. For example at 500°C about 20 hours was required to anneal the single-crystal specimens whereas specimens from the cast ingots would anneal to the same extent in about 1 hour. The multicrystalline ingot would be expected to contain high dislocation density due to the strains introduced on freezing. In addition, the imperfections associated with the grain boundaries in these ingots might also stimulate precipitation.

[10] C. S. Fuller (private communication).

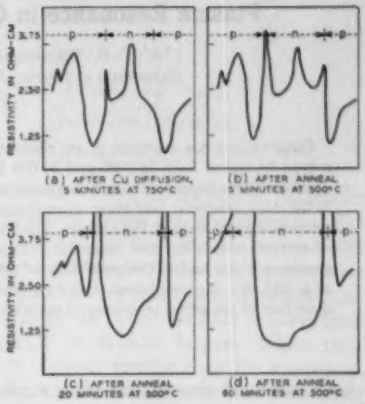


FIG. 3. Resistivity profiles of the copper doped bent rod (R=20 cm) showing stages of anneal at 500°C.

In the determination of copper solubility using resistivity measurements, a considerable variation in acceptor concentration is encountered from sample to sample at any given temperature.[1] It is quite reasonable to suppose that this variation may be largely due to precipitation during the cooling cycle. This point of view is supported by the results obtained with radioactive copper,[1] which measures the total amount of copper in the crystal, including that which precipitates as the crystal is cooled from the high temperature. Solubilities determined by radioactivity have been in general higher than those deduced from resistivity measurement.

As shown in Fig. 3, a copper-doped sample containing 10^6 dislocations/cm^2 annealed completely in about 1 hour at 500°C. However, in the first few minutes of the anneal the acceptor concentration actually increases before the rapid decrease due to precipitation became the dominant effect. This process, which is not understood, is very similar to that observed by Mayburg[11] in his studies of the annealing of quenched germanium at 500°C.

ACKNOWLEDGMENTS

I would like to thank Miss A. D. Mills who measured the resistivity profiles in the deformed samples. It is also a pleasure to acknowledge the many suggestions and informative discussions with W. T. Read, Jr. throughout the course of this work.

[11] S. Mayburg, Phys. Rev. **95**, 38 (1954).

Plasma Resonance in Crystals: Observations and Theory*

G. Dresselhaus, A. F. Kip, and C. Kittel
Department of Physics, University of California, Berkeley, California
(Received June 2, 1955)

Observations are reported of an electron plasma resonance absorption line in a magnetic field in an n-type crystal of InSb at 9000 and 24 000 Mc/sec and at 4° and 77°K. The magnitude of the plasma frequency and its dependence on rf frequency and on the orientation of the figure axis of the specimen relative to the static magnetic field are in approximate agreement with the elementary theory of the plasma resonance developed previously by the authors. The line shape is discussed for experiments carried out under conditions of carrier modulation and magnetic field modulation. A theoretical discussion is given of the effects of specimen shape and of the possibility of detection of minority carrier cyclotron resonance in the presence of a majority carrier plasma. Eddy current effects are discussed, with reference to the possibility of the detection of cyclotron resonance in metals and semimetals.

WE have observed an electron plasma resonance line in a magnetic field in an n-type indium antimonide crystal at microwave frequencies at 4°K and 77°K. Plasma magnetic resonance effects as first observed were discussed and reported briefly by the present authors[1] in a paper on cyclotron resonance in silicon and germanium. In the present paper, we extend the theory in several directions and report the experimental observations in detail. We employ a classical drift-velocity treatment; the results can be extended easily, if required, by using the Boltzmann equation. We believe that the most serious shortcoming of the present treatment is the absence of any treatment of conditions near the surface of the specimen; our discussion is not expected to be valid when either the smallest dimension of the specimen or the skin depth is comparable with the carrier mean free path or the radius of the motion.

I. THEORETICAL

We first review the general nature of a plasma resonance in a magnetic field. This is a simple extension of standard magneto-ionic theory. We consider a spheroidal particle of a crystalline material of dielectric constant ϵ with N conduction electrons per unit volume, the electrons having an isotropic effective mass m^* and an isotropic collision or relaxation frequency ρ $(=\tau^{-1})$. We treat in the present paper only an isotropic effective mass; the electron mass in InSb is known to be isotropic.[2] The particle is placed in a uniform external rf electric field $\mathbf{E}$ perpendicular to the figure or z-axis of the spheroid. The dimensions of the particle are assumed to be small in comparison with the skin depth and the wavelength of the radiation. A static magnetic field $\mathbf{H}$ is applied parallel to the z-axis. The equation of motion of the average or drift velocity of the conduction electrons is given by

$$m^*(d\mathbf{v}/dt+\rho\mathbf{v})=e(\mathbf{E}-L\mathbf{P})+(e/c)\mathbf{v}\times\mathbf{H}, \quad (1)$$

where L is the depolarization factor (analogous to the demagnetization factor) normal to the figure axis of the specimen, and $\mathbf{P}$ is the dielectric polarization. The term $-L\mathbf{P}$ gives the depolarizing electric field associated with the dielectric polarization of the system. We neglect the questionable contribution of a Lorentz or cavity contribution to the local field acting on a carrier; we also neglect the effect of charges induced in the cavity walls by the polarization of the specimen. If χ is the dielectric susceptibility per unit volume of the host crystal, exclusive of carriers, we have for the effective internal electric field acting on the carriers

$$\mathbf{E}_{eff}=\mathbf{E}-L\mathbf{P}=\mathbf{E}-L(\chi\mathbf{E}_{eff}+Ne\mathbf{r}), \quad (2)$$

where N is the conduction electron concentration, and $\mathbf{r}$ is the vector from the center of charge of the ionized donor atoms in the crystal to the center of charge of the conduction electrons. From Eq. (2), we have

$$\mathbf{E}_{eff}=(\mathbf{E}-LNe\mathbf{r})/(1+L\chi). \quad (3)$$

Introducing the notation

$$\mathbf{E}_i=\mathbf{E}/(1+L\chi), \quad (4)$$

$$L_i=L/(1+L\chi), \quad (5)$$

the effective internal electric field becomes

$$\mathbf{E}_{eff}=\mathbf{E}_i-L_iNe\mathbf{r}. \quad (6)$$

The equation of motion in the xy-plane (normal to the figure axis) may be written

$$m^*\left(\frac{d^2\mathbf{r}}{dt^2}+\rho\frac{d\mathbf{r}}{dt}+\frac{L_iNe^2}{m^*}\mathbf{r}\right)=e\mathbf{E}_i+\frac{e}{c}\frac{d\mathbf{r}}{dt}\times\mathbf{H}. \quad (7)$$

The depolarization effect is equivalent to a harmonic restoring force, of force constant L_iNe^2. If $\mathbf{E}_i\sim e^{i\omega t}$, the equation of motion becomes

$$(-\omega^2+\omega_p^2+i\omega\rho)\mathbf{r}=(e/m^*)\mathbf{E}_i+i\omega\mathbf{r}\times\boldsymbol{\omega}_c, \quad (8)$$

* This work has been supported in part by the Office of Naval Research and the U. S. Signal Corps.
[1] Dresselhaus, Kip, and Kittel, Phys. Rev. 98, 368 (1955).
[2] Dresselhaus, Kip, Kittel, and Wagoner, Phys. Rev. 98, 556 (1955).

where $\omega_c = eH/m^*c$, or

$$(i\omega' + \rho)\mathbf{v} = (e/m^*)\mathbf{E}_i + \mathbf{v}\times\omega_c, \tag{9}$$

where

$$\omega_p = (L_i N e^2/m^*)^{\frac{1}{2}} = [LNe^2/m^*(1+L\chi)]^{\frac{1}{2}} \tag{10}$$

is the plasma frequency. We see that the influence of the plasma depolarization may be taken into account by substituting $\omega' = \omega - \omega_p^2\omega^{-1}$ in place of ω in the integrated equations of motion.

A plot of ω_p as a function of the depolarization factor L and of the axial ratio of a spheroid is given in Fig. 1, taking for convenience $\chi = 1$ which corresponds to $\epsilon = 13.6$, a value not unrepresentative of several classes of semiconductors. The shape dependence of the plasma frequency for this value of the dielectric susceptibility is not very marked, except for quite flat geometries.

The rate of energy dissipation per unit volume is given by

$$\mathcal{P} = \tfrac{1}{2}\,\mathrm{Re}[\mathbf{E}\cdot(d\mathbf{P}^*/dt)]. \tag{11}$$

Now

$$\mathbf{P} = \chi\mathbf{E}_{\mathrm{eff}} + Ne\mathbf{r}, \tag{12}$$

or, using Eq. (3),

$$\frac{d\mathbf{P}}{dt} = \frac{i\omega\chi}{1+L\chi}\mathbf{E} + \frac{1}{1+L\chi}Ne\mathbf{v}.$$

If a tensor conductivity component, σ_{xx} is defined in terms of the solution of Eq. (9) by

$$\sigma_{xx} = Nev_x/E_i^x,$$

we have, for the applied electric field in the x-direction,

$$\mathcal{P} = \frac{\frac{1}{2}|E^x|^2}{(1+L\chi)^2}\,\mathrm{Re}(\sigma_{xx}) = \tfrac{1}{2}|E_i^x|^2\,\mathrm{Re}(\sigma_{xx}). \tag{13}$$

The same expression is obtained on calculating $\mathcal{P}$ as $\frac{1}{2}\,\mathrm{Re}(\mathbf{E}_i\cdot\mathbf{j}^*)$, where $\mathbf{j}$ is the current density. We may note that $\omega_p^2 = L_i\sigma\rho$, where ρ is the relaxation frequency.

The real part σ_R of the conductivity σ_{xx} is given by

$$\frac{\sigma_R}{\sigma_0} = \frac{1+\nu'^2+\nu_e^2}{(1+\nu_e^2-\nu'^2)^2+4\nu'^2}, \tag{14}$$

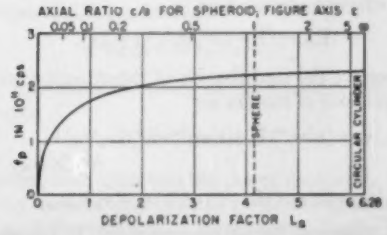


FIG. 1. Plasma frequency $f_p = \omega_p/2\pi$ for $m^* = 0.01m$, $\chi = 1$, and $N = 10^{24}$ carriers/cm³, calculated as a function of depolarization factor L_a.

where

$$\sigma_0 = Ne^2\tau/m^*;$$
$$\nu = \omega\tau; \quad \nu_p = \omega_p\tau;$$
$$\nu_e = \omega_e\tau = (eH/m^*c)\tau;$$
$$\nu' = (\omega - \omega_p^2\omega^{-1})\tau = \nu - \nu_p^2\nu^{-1}.$$

The variation of σ_R with the ratio of plasma frequency to experimental frequency is plotted in Fig. 2 for several values of the line sharpness parameter $\omega\tau = \nu$. When $\omega_p = \omega$ the conductivity is a maximum and equal to the value at zero frequency. In the absence of plasma effects ($\omega_p = 0$) the rf conductivity is lower than the dc conductivity, as is well known. When the plasma frequency is much greater than the experimental frequency ($\omega_p \gg \omega$), the electron motion is stiffened by the plasma effects, and the conductivity is reduced; to the limit $\sigma_R \cong \omega^2/L_i^2\sigma_0$. In this limit the resonance frequency is independent of the effective mass and is $\omega \cong L_i Nec/H$. The reduction in rf losses at high conductivities was pointed out in connection with some earlier experimental observations.[3]

We should note that plasma effects may be expected to have an important influence on microwave conductivity measurements on semiconductors. In such experiments the geometry is usually planned with the hope that the wave guide or cavity walls will act effectively to short-circuit the depolarization fields, so plasma effects may be neglected ($L = 0$). It has not been demonstrated theoretically or experimentally to our knowledge that the assumption $L = 0$ is sufficiently valid for all of the diverse experimental arrangements which have been employed. We suggest it is possible that plasma effects may be responsible in part for the discrepant results reported by various observers for the microwave conductivity of germanium.

The development of a magnetoplasma resonance from a cyclotron resonance as the carrier concentration (plasma frequency) is increased is illustrated in Fig. 3, where the conductivity is plotted against the magnetic field intensity, both in normalized units. The relaxation

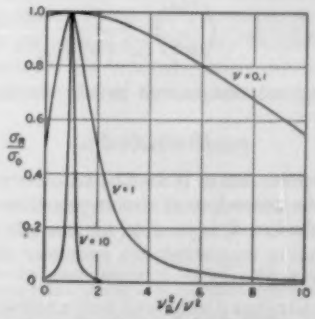


FIG. 2. Calculated variation of σ_R in zero magnetic field, as a function of $(\nu_p^2/\nu') = (\omega_p^2/\omega')$, for several values of $\nu = \omega\tau$.

[3] Portis, Kip, Kittel, and Brattain, Phys. Rev. 90, 988 (1953).

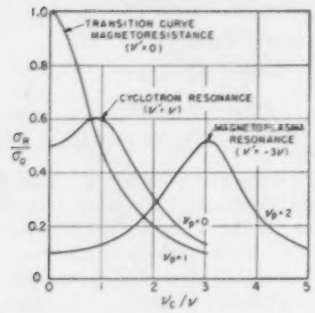


FIG. 3. Development of magnetoplasma resonance; theoretical plot of effective conductivity σ_R vs $\omega_c/\omega=\nu_c/\nu$ for $\nu=1$ and several values of ν_p.

time is kept constant with $\nu=1$. At zero plasma frequency we have the curve $\nu'=\nu$ for which a cyclotron resonance is just resolved For $\nu_p=\nu$ we have the transition curve $\nu'=0$ which is of the form of an ordinary dc magnetoresistance effect. For $\nu_p=2\nu$ we have the curve $\nu'=-3\nu$ which exhibits a distinct magnetoplasma resonance.

The striking consequences for the magnetoplasma resonance of a decrease in the experimental frequency are shown by Fig. 4. The two curves are drawn for a constant relaxation time and constant plasma frequency ($\nu_p=2$), but for two experimental frequencies $\nu=1.0$ and $\nu=0.5$. It is seen that a *reduction* in frequency by a factor of two almost *doubles* the actual value of the magnetic field for which the maximum of the magnetoplasma resonance occurs.

It is sometimes convenient to work with the components of the conductivity tensor. We write

$$j_x=\sigma_{xx}E_x+\sigma_{xy}E_y; \quad j_y=\sigma_{yx}E_x+\sigma_{yy}E_y. \tag{15}$$

For the problem considered above,

$$\sigma=\sigma_{xx}=\sigma_{yy}=\left(\frac{Ne^2}{m^*}\right)\frac{i\omega'+\rho}{(i\omega'+\rho)^2+\omega_c^2}; \tag{16}$$

$$\sigma_{xy}=-\sigma_{yx}=\left(\frac{Ne^2}{m^*}\right)\frac{\omega_c}{(i\omega'+\rho)^2+\omega_c^2}. \tag{17}$$

The off-diagonal components satisfy the reciprocity relation

$$\sigma_{xy}(H)=\sigma_{yx}(-H).$$

The tensor formulation is useful particularly in treatments of the nonreciprocal circuit properties of semiconductor devices. It is possible, for example, to utilize the cyclotron or magnetoplasma resonance effects in a semiconductor gyrator.

Elliptical Cross Section

If the depolarization factors L^x, L^y in the xy plane—the plane of the cyclotron motion—are not equal, we are to replace L_i in Eq. (10) for the plasma frequency by an appropriate function. We have

$$\begin{aligned} E_{\text{eff}}{}^x&=(E^x-L^xNex)/(1+L^x\chi);\\ E_{\text{eff}}{}^y&=(E^y-L^yNey)/(1+L^y\chi). \end{aligned} \tag{18}$$

The determinantal equation for free undamped oscillations is

$$\begin{vmatrix} i(\omega-\omega_{px}{}^2\omega^{-1}) & -\omega_c \\ \omega_c & i(\omega-\omega_{py}{}^2\omega^{-1}) \end{vmatrix}=0, \tag{19}$$

so resonance occurs when

$$\omega_c{}^2=(\omega-\omega_{px}{}^2\omega^{-1})(\omega-\omega_{py}{}^2\omega^{-1})=\omega_x'\omega_y', \tag{20}$$

where $\omega_{px}{}^2=L_i{}^xNe^2/m^*$; $\omega_{py}{}^2=L_i{}^yNe^2/m^*$; $L_i{}^x=L^x/(1+L^x\chi)$; $L_i{}^y=L^y/(1+L^y\chi)$. When $\omega_{px}{}^2$, $\omega_{py}{}^2\gg\omega^2$, we have

$$\omega_c\cong(\omega_{px}\omega_{py}/\omega)=(L_i{}^xL_i{}^y)^{\frac{1}{2}}Ne^2/m^*\omega, \tag{21}$$

so when plasma effects are dominant the magnetic field for magnetoplasma resonance is determined by the square root of the product of the effective depolarization factors $L_i{}^x$, $L_i{}^y$.

If one depolarization factor, say $L_i{}^y$, is zero, the apparent cyclotron frequency is given by $\omega_c{}^2=\omega^2-\omega_{px}{}^2$. This shows that when one of the plasma frequencies (ω_{py}) is small, a *resonance* may only be obtained when the other (ω_{px}) is also small, in order that ω^2 may be larger than $\omega_{px}{}^2$.

Minority Carriers

We wish now to study the behavior of a low concentration N_s of minority carriers (which, for example, may be injected optically) in the presence of a high concentration N_h ($\omega_p\gg\omega$) of majority carriers h. In particular, we wish to determine the effect of the majority carriers on the cyclotron resonance of the minority carriers. In our work on cyclotron resonance in indium antimonide we observed the cyclotron resonance of a low concentration (possibly 10^4-10^5 carriers/cm^3) of high-mobility electrons and holes produced by optical excitation against the background of a relatively high concentration ($\sim 4\times10^{14}$ per cm^3) of low-mobility holes, presumably in an impurity band.[4] The present calculation examines the possible interaction effects in such a system.

The effective electric field is

$$\mathbf{E}_{\text{eff}}=(\mathbf{E}-LN_h\epsilon\mathbf{r}_h)/(1+L\chi), \tag{22}$$

if we neglect the contribution of the minority carriers. The equations of motion are

$$(i\omega+\rho_s)\mathbf{v}_s=(e/m_s^*)\mathbf{E}_s+i(m_h^*/m_s^*)(\omega_p{}^2/\omega)\mathbf{v}_h+\mathbf{v}_s\times\boldsymbol{\omega}_{cs}; \tag{23}$$

$$(i\omega+\rho_h)\mathbf{v}_h=(e/m_h^*)\mathbf{E}_s+i(\omega_p{}^2/\omega)\mathbf{v}_h+\mathbf{v}_h\times\boldsymbol{\omega}_{ch}, \tag{24}$$

[4] H. J. Hrostowski (private communication); H. Fritzsche and K. Lark-Horovitz (private communication). We are indebted to all three workers for helpful communication of their unpublished results.

where $\omega_p^2=L_i N_h e^2/m_h^*$. We note that in our approximation the motion of the majority carriers is independent of the motion of the minority carriers. Thus, if **E** is in the x-direction,

$$v_{hx}=\frac{(e/m_h^*)E_i^x(i\omega'+\rho_h)}{(i\omega'+\rho_h)^2+\omega_{ch}^2}; \tag{25}$$

$$v_{hy}=-\frac{(e/m_h^*)E_i^x\omega_{ch}}{(i\omega'+\rho_h)^2+\omega_{ch}^2}. \tag{26}$$

In this degree of generality, the equations of motion for the minority carriers are rather tedious and unrevealing. We examine first the equations in the approximation $\omega_{ch}\ll\rho_h$, ω'; that is, magnetic effects on the majority carriers are neglected. We have

$$(i\omega+\rho_e)v_{ex}-\omega_{ce}v_{ey}=(eE_i^x/m_e^*)\left[1+\frac{i\omega_p^2}{(i\omega'+\rho_h)\omega}\right]; \tag{27}$$

$$(i\omega+\rho_e)v_{ey}+\omega_{ce}v_{ex}=0; \tag{28}$$

so

$$\frac{v_{ex}}{E_i^x}=\left(\frac{e}{m_e^*}\right)\frac{(i\omega+\rho_h)(i\omega+\rho_e)}{[(i\omega+\rho_e)^2+\omega_{ce}^2](i\omega'+\rho_h)}. \tag{29}$$

The total current density is

$$j_x=N_e e v_{ex}+N_h e v_{hx}. \tag{30}$$

If we set

$$\sigma_{Rh}+i\sigma_{Ih}=\frac{N_h e^2/m_h^*}{i\omega'+\rho_h}, \tag{31}$$

$$\sigma_{Re}+i\sigma_{Ie}=\frac{(N_e e^2/m_e^*)(i\omega+\rho_e)}{(i\omega+\rho_e)^2+\omega_{ce}^2}, \tag{32}$$

we have for the total conductivity $\sigma(=j_x/E_i^x)$:

$$\sigma=\sigma_{Rh}+i\sigma_{Ih}+(\sigma_{Re}+i\sigma_{Ie})\frac{i\omega+\rho_h}{i\omega'+\rho_h}, \tag{33}$$

so the power dissipation associated with the minority carriers is determined by

$$\mathrm{Re}\Delta\sigma=\mathrm{Re}(\sigma-\sigma_{Rh})=\frac{\sigma_{Re}(\omega\omega'+\rho_h^2)+\sigma_{Ie}(\omega'-\omega)\rho_h}{\omega'^2+\rho_h^2}. \tag{34}$$

If plasma effects are dominant

$$\mathrm{Re}\Delta\sigma\approx-\sigma_{Re}(\omega/\omega_p)^2. \tag{35}$$

If the relaxation frequency of the majority carriers is dominant,

$$\mathrm{Re}\Delta\sigma=\sigma_{Re}. \tag{36}$$

In intermediate situations the power dissipation by the minority carriers may contain a mixture of absorptive (σ_{Re}) and dispersive (σ_{Ie}) components. In the p-type indium antimonide crystal produced by the Bell Telephone Laboratories and used for our cyclotron resonance work, we have approximately, at 4°K, $N_h\approx 4\times10^{14}$/cm³ and $\rho_h\approx 2\times10^{14}$ sec^{-1} as calculated from the conductivity $\sigma_h=4\times10^{-4}$ (ohm-cm)$^{-1}$ and an effective mass[5] of m; taking $L_i\approx 1$, we have $\omega_{ph}\approx 3\times10^{11}$ sec^{-1}. We see therefore that at the experimental frequency $\omega=1.5\times10^{11}$ the relaxation frequency may be dominant in this specimen, and the energy absorption by minority carriers probably is not affected by the presence of the majority carriers.

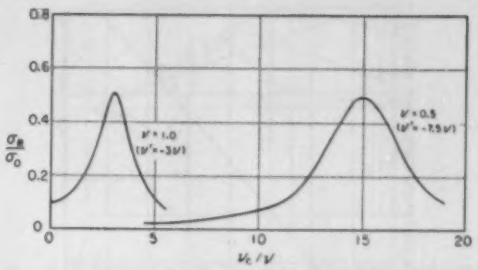


FIG. 4. Calculated effect of reducing experimental frequency by a factor of two. Plot of conductivity vs $\omega_c/\omega=\nu_c/\nu$ for $\nu=1.0$ and $\nu=0.5$, keeping $\nu_p=2$.

Cyclotron and Plasma Resonance Under Eddy Current Conditions

We calculate now the magnetic field dependence of the rate of energy absorption in an infinite slab of material thick in comparison with the rf skin depth.[6] Our primary interest is in the shape of the absorption line under eddy current conditions. We assume explicitly that the carrier mean free path is small in comparison with the skin depth. This assumption limits the region of validity of the treatment to moderately low carrier concentrations. Two geometries are of interest, one with the static magnetic field parallel to the plane of the slab and the second with the magnetic field normal to the plane. We take the plane of the slab to be the xy plane. The results suggest that under eddy current conditions and the above assumptions it is not possible to obtain a distinctive cyclotron resonance line. With carriers produced by optical absorption it should be possible to get around the difficulty for the perpendicular geometry if the injected carriers can be localized in the surface region of the specimen, as by recombination.

Parallel Geometry

The static magnetic field is in the y-direction. The equations to be solved are the cyclotron equations (15) through (17), and the wave equation in a conducting

[5] The effective mass should be that of holes in the impurity band; we have taken this as m in complete ignorance of what to do. It may be that a much higher mass is applicable; in the limit of high mass the majority carriers will have little effect on the cyclotron resonance of the minority carriers.

[6] See also B. Donovan and E. H. Sondheimer, Proc. Phys. Soc. (London) **A66**, 849 (1953).

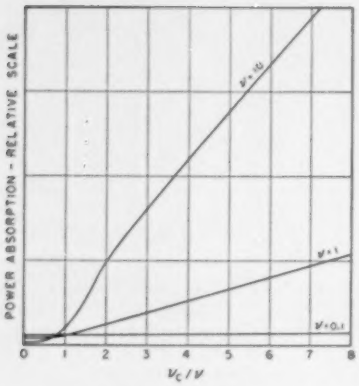


FIG. 5. Calculated power absorption *vs* magnetic field intensity under eddy current conditions with H parallel to the surface.

medium

$$\nabla^2\mathbf{E}=\frac{4\pi}{c^2}\frac{\partial\mathbf{j}}{\partial t}+\frac{\epsilon}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2}. \tag{37}$$

The right-hand side of this equation is zero for the z component. We assume E_x has a solution of the form $e^{i(\omega t-kz)}$; the wave equation gives us

$$k^2=-i\frac{4\pi\omega\sigma}{c^2}+\frac{\epsilon\omega^2}{c^2}. \tag{38}$$

With the appropriate restrictions this expression gives the usual low frequency eddy current result.

The electromagnetic behavior of the material is described conveniently by the intrinsic surface impedance

$$Z=\frac{E_x}{H_y}\bigg|_{z=0}=-\frac{i\omega}{c}\frac{E_x}{\partial E_x/\partial z}=\frac{\omega}{ck}. \tag{39}$$

The rate of energy loss per unit area is

$$S=(c/8\pi)\,|H_y|^2\,\mathrm{Re}(Z), \tag{40}$$

from the Poynting vector. If we may neglect $\epsilon\omega^2/c^2$ in (38), as is often justified, we have that the shape of the absorption line is determined by

$$\mathrm{Re}(Z)\propto\mathrm{Re}(i/\sigma)^{\frac{1}{2}}=\left[\frac{(\sigma_R^2+\sigma_I^2)^{\frac{1}{2}}+\sigma_I}{2(\sigma_R^2+\sigma_I^2)}\right]^{\frac{1}{2}}, \tag{41}$$

where

$$\sigma=\sigma_R+i\sigma_I. \tag{42}$$

The unusual line shape is shown in Fig. 5 for a pure cyclotron resonance, with no specific plasma effects—that is, the ratio of the skin depth to the breadth of the specimen is taken to be vanishingly small. There is no recognizable cyclotron resonance line.

Perpendicular Geometry

The static magnetic field is in the z-direction. From the electromagnetic wave equation

$$\begin{aligned}-k^2E_x&=i(4\pi\omega/c^2)(\sigma E_x+\sigma_{xy}E_y)-(\epsilon\omega^2/c^2)E_x;\\ -k^2E_y&=i(4\pi\omega/c^2)(\sigma_{yx}E_x+\sigma E_y)-(\epsilon\omega^2/c^2)E_y.\end{aligned} \tag{43}$$

We obtain k from the following determinantal equation, where $p=4\pi\omega/c^2$:

$$\begin{vmatrix} k^2+ip\sigma-\dfrac{\epsilon\omega^2}{c^2} & ip\sigma_{xy} \\ -ip\sigma_{xy} & k^2+ip\sigma-\dfrac{\epsilon\omega^2}{c^2}\end{vmatrix}=0, \tag{44}$$

so

$$k^2=(\epsilon\omega^2/c^2)-ip\sigma\pm p\sigma_{xy}, \tag{45}$$

and

$$E_x=\pm iE_y. \tag{46}$$

If we enforce the condition that the radiation field at the surface shall be linearly polarized, the field components are given by

$$\begin{aligned}E_x&=\tfrac{1}{2}E_0(e^{-ik_+z}+e^{-ik_-z});\\ iE_y&=\tfrac{1}{2}E_0(e^{-ik_+z}-e^{-ik_-z}).\end{aligned} \tag{47}$$

Thus, $H_y=(ic/\omega)(\partial E_x/\partial y)=(c/2\omega)E_0(k_++k_-)$ at the surface, and, neglecting $\epsilon\omega^2/c^2$,

$$\mathrm{Re}(Z)\propto\mathrm{Re}\left\{\left(\frac{i\omega}{4\pi\sigma_0}\right)^{\frac{1}{2}}\times\frac{(1+i\omega'\tau-i\omega_c\tau)^{\frac{1}{2}}(1+i\omega'\tau+i\omega_c\tau)^{\frac{1}{2}}}{(1+i\omega'\tau-i\omega_c\tau)^{\frac{1}{2}}+(1+i\omega'\tau+i\omega_c\tau)^{\frac{1}{2}}}\right\}. \tag{48}$$

A plot of the power absorption as a function of the magnetic field intensity is given in Fig. 6. There is no recognizable cyclotron resonance line.

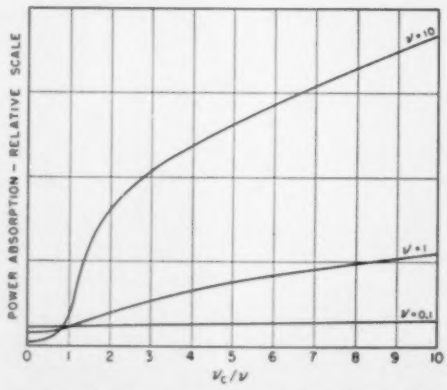


FIG. 6. Calculated power absorption *vs* magnetic field intensity under eddy current conditions with H normal to the surface.

Line Shape

It is often convenient in microwave resonance experiments to present as the output signal the modulated component of the rf power absorption in the specimen. The modulation in our experiments may be produced by modulating the carrier concentration,[7] which may be accomplished conveniently by modulating a light source which produces internal photoelectrons in the specimen, or by modulating the magnetic field intensity, a standard method in electron and nuclear spin resonance experiments.

In carrier modulation, the signal S is proportional to $\partial\sigma_R/\partial N$, where σ_R is given by Eq. (14) and N is the carrier concentration. We find that

$$S \propto \mathrm{Re}\left[\sigma - \frac{\nu_p^2}{\nu}\frac{\partial\sigma}{\partial\nu'}\right], \tag{49}$$

or

$$S \propto \mathrm{Re}\left\{\frac{1+i\nu'}{(1+i\nu')^2+\nu_c^2} - \frac{i\nu_p^2[\nu_c^2-(1+i\nu')^2]}{\nu[\nu_c^2+(1+i\nu')^2]^2}\right\}. \tag{50}$$

If $\nu' \gg \nu$

$$S \propto \left\{\frac{1+\nu_c^2-\nu'^2}{(1+\nu'^2-\nu_c^2)^2+4\nu'^2} + 8\nu'^2\nu_c^2\frac{(1+\nu_c^2-\nu'^2)}{[(1+\nu'^2-\nu_c^2)^2+4\nu'^2]^2}\right\} \tag{51}$$

and the phase of the signal changes when

$$\nu_c^2 = \nu'^2 - 1. \tag{52}$$

A plot of calculated line shapes is given in Fig. 7, in the approximation $\nu' \gg \nu$. The line shapes are quite distinctive and are somewhat reminiscent of the shapes of ordinary spin resonance absorption lines under eddy current conditions. We note that if plasma effects may be neglected the signal is simply of the form of σ_R, an absorption shape. It should therefore be possible to distinguish in most instances plasma resonances from cyclotron resonances by the line shape.

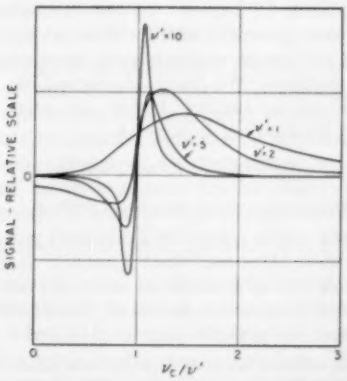


FIG. 7. Plot of theoretical line shape *vs* magnetic field when the carrier concentration is modulated and the detected signal is the modulated component of the power absorption. The plot is for the limiting case $\nu \ll \nu_p$, so $\nu' \cong -\nu_p^2/\nu$.

[7] A. F. Kip, Physica **20**, 813 (1954).

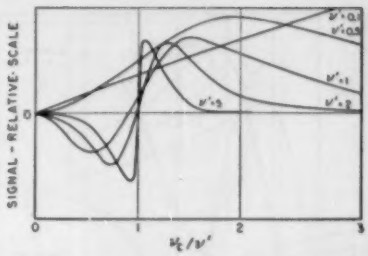


FIG. 8. Plot of theoretical line shape *vs* magnetic field when the magnetic field is modulated and the detected signal is the modulated component of the power absorption.

With magnetic field modulation the signal S is of the form

$$S \propto \partial\sigma_R/\partial H, \tag{53}$$

or

$$S \propto \nu_c\left\{\frac{(1+\nu_c^2-\nu'^2)^2+4\nu'^2(\nu_c^2-\nu'^2)}{[(1-\nu'^2+\nu_c^2)^2+4\nu'^2]^2}\right\}. \tag{54}$$

The calculated line shapes are plotted in Fig. 8; it is seen that the lines tend to be more antisymmetric about the resonance frequency than was true with carrier modulation.

II. EXPERIMENTAL RESULTS

Although the plasma resonance absorption line is in general relatively strong, it is not entirely a trivial matter to arrange to observe the resonance experimentally. It is not convenient to sweep in frequency over a wide range, so we have preferred to use a magnetic field to sweep the plasma frequency through the experimental frequency. It is desirable to use a crystal in which the carrier concentration is independent of temperature,[8] so the heating or ionization effects of the microwave field will not change the carrier concentration. We were fortunate enough to obtain from the Battelle Memorial Institute, through the kindness of Dr. A. C. Beer, an n-type indium antimonide crystal which is very well suited to our purpose. The crystal is described by Dr. Beer as having an average extrinsic carrier density of 1.4×10^{14} cm^{-3}; we are inclined to believe on the basis of our plasma measurements that the fragment used in our work actually had a carrier density of $(0.5\pm0.2)\times10^{14}$ cm^{-3}. The electron density at somewhat higher concentrations in n-InSb is known

[8] There is some interest in experiments on materials in which the carrier concentration can be varied by varying the temperature, and our original observation of a plasma resonance in germanium was made in this way, but we have not used the method in a systematic way.

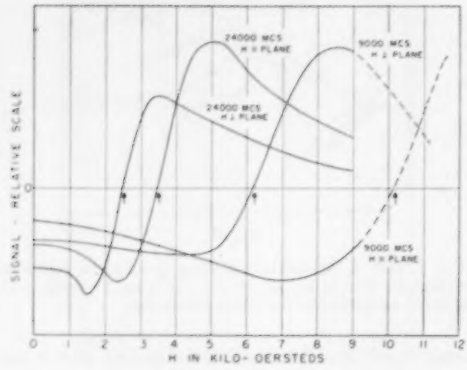


FIG. 9. Experimental plasma resonance absorption signals obtained with carrier modulation in a thin disk of *n*-type InSb at 77°K, at 9000 Mc/sec and 24 000 Mc/sec as indicated. The static magnetic field is directed normal to or parallel to the plane of the disk in separate runs. The broken lines connect the curves below 9000 oersteds with single terminal points determined at higher fields. The resonance condition is determined approximately by the crossover points.

to be independent of temperature[9] between the temperatures of liquid helium and liquid nitrogen. In our sample the plasma frequency did not appear to shift between these two temperatures, so we believe with some confidence that the electron concentration is indeed approximately constant. At high rf power levels at *X*-band at 77°K it appeared to be possible to alter the carrier concentration, however.

It is desired that the carrier concentration and effective mass correspond approximately to a convenient experimental frequency. As shown roughly in Fig. 1, the properties of our specimen put the plasma frequency close enough to the microwave range, so a reasonable laboratory magnetic field acting on the light effective mass of the carrier makes possible measurements at both *X*-band and *K*-band. It is also necessary that the relaxation time of the carriers be long enough to permit the resonance to be resolved distinctly. The mobility measured at Battelle at 80°K is reported to be about 500 000 cm²/volt-sec. Taking the effective mass[10] to be $m^*=0.013m$, we have $\tau=\mu m^*/e\cong 4\times 10^{-12}$ sec. As at *K*-band we obtained plasma resonance at a cyclotron angular frequency of about 4×10^{12} radians/sec, we should expect a resolution $\omega_c\tau$ of about 16, more than adequate. The actual resolution was only about 2; we believe that the actual resolution is probably limited by unavoidable irregularities in the shape of the test specimen. The fact that the resolution did not improve on moving to *X*-band[11] supports our opinion.

It is required further that the thickness of the specimen be less than the rf skin depth. Based on the reported mobility $\mu=5\times 10^5$ cm²/volt-sec and carrier concentration $N=1.4\times 10^{14}$ cm^{-3}, we calculate that the skin depth at *K*-band is about 0.1 mm. The value may be raised to 0.2 mm if the correct *N* should be 0.5×10^{14} cm^{-3}. The skin depth is too small, taken in combination with the poor mechanical properties of the material, for us to fabricate easily a suitable well-formed spheroidal test specimen. Through persistent efforts, Dr. Lawrence Hadley was able to isolate an irregular disk roughly 0.4 mm×0.4 mm×0.1 mm. This specimen was good enough to enable us to confirm approximately the principal predictions of the magnetoplasma theory, but we can claim none of the high accuracy and reproducibility often associated with other types of resonance experiments. We also checked qualitatively the general features of the results using a second specimen in the form of a short thin needle, making runs at 77°K and 4°K, with no significant differences at the two temperatures.

The experimental absorption curves obtained with the disk-like specimen are plotted in Fig. 9, for runs taken at 77°K under carrier (light) modulation.[12] The observed curves show the general features of the calculated curves, Fig. 7, but seem to be smudged out as might be expected if the depolarization factors vary somewhat from point to point within the irregular specimen. The resonance field at 24 000 Mc/sec for the normal field orientation is apparently near 2500 oersteds, as compared with 3500 oersteds in the parallel field orientation.

We note that in our specimen the effect of the plasma may be described as shifting the position of the cyclotron resonance to higher magnetic fields; the shift is by a factor of 20 or more at *K*-band and 200 or more at *X*-band.

The ratio of 0.7 between the two orientations is in approximate agreement with the ratio calculated using Eq. (21) and the approximate axial dimensions *a*, *b*, *c* of the specimen. We have $(L_a/4\pi)=0.70$; $(L_b/4\pi)=(L_c/4\pi)=0.15$; further, taking the value of the dielectric constant[13] of InSb as 14,

$$L_i^a=0.87;\quad L_i^b=L_i^c=0.64, \tag{55}$$

so the calculated ratio of resonance fields is

$$H(\perp)/H(\parallel)=0.64/[(0.64)(0.87)]^{\frac{1}{2}}=0.86, \tag{56}$$

with an appreciable range of error because of the uncertain depolarization factors of the specimen. The thickness actually varies between 0.07 and 0.12 mm;

[9] H. J. Hrostowski (private communication).

[10] See reference 2. We take advantage of this occasion to correct an error in the reference cited: the electron spin resonance reported for InSb is now believed to be associated instead with both cavities used in the work. It has been necessary to silver-plate the cavities to eliminate the spurious signal. The specimen reported as *n*-type in the reference is *p*-type extrinsic at low temperatures.

[11] We recall that decreasing ω increases ω_c, according to Eq. (14).

[12] The experimental methods employed were similar in general to those described in reference 1; in the plasma work some difficulty is caused by the high loss of the specimen.

[13] Avery, Goodwin, Lawson, and Moss, Proc. Phys. Soc. (London) **B67**, 761 (1954). Dr. Beer has called our attention to several unpublished results suggesting that the dielectric constant may be close to 17. If we use this value, Eq. (58) would give *N* close to 0.5×10^{14} cm^{-3}, which might be increased to 0.6×10^{14} cm^{-3} by the consideration which follows Eq. (58).

further, the injected carriers may not cover the specimen uniformly. The observed resonance field ratio 0.7 is somewhat low.

We shall now estimate the carrier concentration N from the K-band data in the normal field orientation. The approximate resonance condition is, from Eq. (14),

$$-\omega\omega_c \cong \omega_p^2 = L_i N e^2/m^*. \tag{57}$$

Here $\omega = 1.5\times10^{11}$ radians/sec; $\omega_c = eH/m^*c \cong 3\times10^{12}$ radians/sec for $H = 2500$ oersteds and $m^* = 0.013m$; and $L_i^e = 0.64$ from Eq. (55). Thus

$$N \cong m^*\omega\omega_c/L_i e^2 \cong 0.4\times10^{14}\ \text{cm}^{-3}, \tag{58}$$

as compared with the electron concentration of 1.4×10^{14} cm^{-3} reported by Dr. Beer for a part of the crystal as grown. We do not believe the discrepancy necessarily is significant in view of the distribution in impurity concentration known to occur in other semiconductor crystals. The anisotropy discrepancy suggests that we have overestimated L_i^e, and a correction for this would make the observed $N \cong 0.5\times10^{14}$ cm^{-3}.

Experimental results are shown in Fig. 10 for a second specimen which had a better geometry than that used in Fig. 9; the improvement in resolution is evident.

It is a remarkable feature of the plasma resonance condition (57) that the resonance moves to higher magnetic fields as the frequency is lowered. The frequency ratio 9000 Mc/sec/24 000 Mc/sec is 1/2.7; the corresponding resonance field ratio is expected to be 2.7. The observed increase on going from K-band to X-band is from 2500 oersteds to 6200 oersteds in the perpendicular orientation, a factor of 2.5. In the parallel orientation the results are less reliable because of limitations on the available magnetic field, but the corresponding factor appears to be about 2.9. The effects of frequency are thus in general agreement with expectation.

III. SUMMARY

We have given an elementary and simplified theory of plasma resonance in a magnetic field in semiconductor crystals for carriers having a single isotropic effective mass. The dependence of the magnetic field at resonance on microwave frequency, specimen shape and orientation, and carrier concentration has been confirmed approximately by observations on conduction electrons in indium antimonide. The plasma effect, apart from any intrinsic interest, is important in the interpretation of cyclotron resonance experiments at high carrier concentrations; that is, at concentrations over perhaps 10^{13} cm^{-3}. It remains to be seen whether the depolarizing electric fields which give rise to the plasma frequency can be short-circuited by suitable contact between the specimen and the walls of the resonant cavity. It may not be possible to carry out cyclotron resonance experiments at high carrier densities unless the plasma frequency can be suppressed. In normal metals the plasma frequency for $L = 4\pi$ is in the near ultraviolet spectral region.

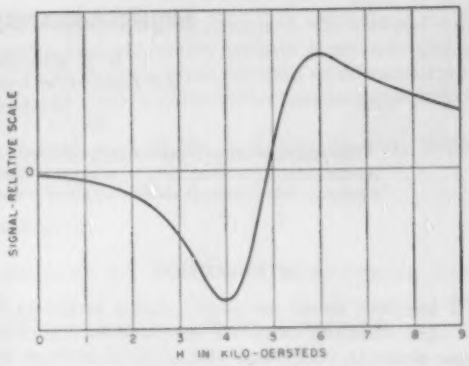


FIG. 10. Experimental plasma resonance absorption signal obtained with carrier modulation in a thin rectangular specimen ($a = 0.04$ cm, $b = 0.026$ cm, and $c = 0.008$ cm) of n-type InSb at 77°K and 24 000 mc. The magnetic field is along the c-axis. The position of the resonance corresponds to an electron charge concentration of $\sim 0.8\times10^{14}$/cm^3. The width indicates an apparent relaxation time of about $\omega_c\tau \gtrsim 4$. The improvement in resolution with respect to the curves shown in Fig. 9 is attributed to the smoother geometry of the present specimen.

It may be possible that the properties of the electron plasma in indium antimonide may find application in oscillators in the millimeter and submillimeter region. It is relatively easy to obtain a pure plasma resonance in this region, and the Q of the resonance may be of the order of 50 to 100 at liquid nitrogen temperature. The oscillator might be tuned by means of a magnetic field.

Apart from the overriding importance of depolarizing fields in metals, we note from the calculations presented above that the specimen must, regardless of carrier concentration, be thinner than a skin depth if the plot of power absorption *vs* magnetic field is to exhibit a resonance maximum. This point is not considered correctly in Dingle's discussion[14] of diamagnetic resonance in metals, and we have at present no reason to believe that resonance absorption can occur under the conditions envisaged in his paper. This remark also applies to the reference by Dorfman[15] to cyclotron resonance in metals.

We are indebted to Mr. G. Wagoner for assistance in the measurements. The crystal of indium antimonide used in the work was supplied by the Battelle Memorial Institute through the kindness of Dr. A. C. Beer. It was developed as part of contractual work undertaken for Wright Air Development Center, U. S. Air Force. Dr. Lawrence Hadley, on leave from Dartmouth College, kindly undertook the difficult preparation of the specimen. We wish also to thank Mr. R. Behringer for checking the manuscript. We wish to thank Professor L. B. Loeb and Dr. D. Judd for helpful conversations.

[14] R. B. Dingle, Proc. Roy. Soc. (London) **A212**, 38 (1952); Phys. Rev. **98**, 550 (1955).
[15] J. G. Dorfman, Doklady Akad. Nauk (S.S.S.R.) **81**, 765 (1951); Phys. Rev. **96**, 1704 (1954).

Superconductivity of Zirconium Alloys

B. T. Matthias and E. Corenzwit
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received June 17, 1955)

The superconducting transition temperatures of alloys between zirconium and VIIIth column elements are substantially higher than those of zirconium. The same is true for zirconium-gold alloys. This observation confirms a rule previously established about the height of transition temperatures.

I. INTRODUCTION

IT has been shown previously[1] that a maximum in superconducting transition temperatures occurs for those elements, compounds and alloys which have an average valence electron per atom ratio near 5 or 7. This behavior seems analogous to that of γ, the electronic term of the specific heat,[2] which shows a maximum for elements with 3, 5, or 7 valence electrons per atom. It seemed likely, therefore, that a maximum in the superconducting transition temperature should occur simultaneously with a maximum for γ. Elements with 4 valence electrons per atom should have, therefore, minimum transition temperatures. In fact, hafnium is the lowest known superconductor at 0.32°K. The purpose of this note is to show data on zirconium alloys with electron concentrations on both sides of 4 and which show the expected minimum in the superconducting transition temperature.

II. THE VARIATION OF ELECTRON DENSITIES

Zirconium, which also has 4 valence electrons per atom, becomes superconducting at 0.75°K. Zr-Rh alloys show rather high transition temperatures,[1] presumably due to the increase in electron density. This same effect should now be shown in varying degrees by alloying Zr with the other elements in the VIIIth column of the periodic system. The results confirming this assumption are given in Table I.

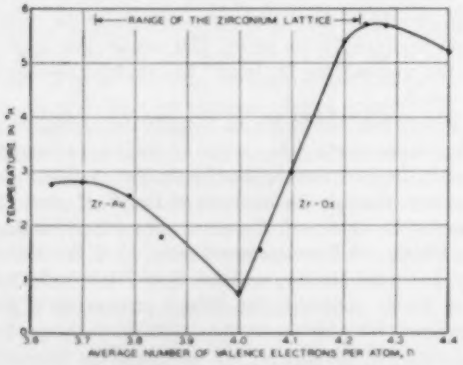


FIG. 1. Superconducting transition temperatures as function of electron concentration.

Table I shows the superconducting transition temperatures of alloys between Zr and 10% (atomic) of the given element. As can be seen, they all raise the transition temperature of zirconium.

In order to demonstrate the existence of a minimum, one had to find superconducting Zr alloys with an average valence electron per atom concentration *below* 4. These were finally discovered in the zirconium-gold system. For about 10% (atomic) Au the transition temperature is at a maximum of 2.80°K.

In Fig. 1, the variation of transition temperature with varying electron concentration is shown. To illustrate the rise in temperature for an increase of electrons, osmium alloys were chosen. Any other element in the VIIIth column would have exhibited the same behavior qualitatively. We are aware that by comparing zirconium-gold alloys on the left side to zirconium-osmium alloys on the right side, we are stretching the analogy. Our excuse is that within the indicated region we preserve essentially the same Zr lattice in both alloys. The minimum transition temperature of Zr is apparent. The maxima, however, on both sides are not those mentioned earlier and are only due to the onset of a second phase.

III. Zr ALLOYS WITH FERROMAGNETIC ELEMENTS

It seems worthwhile to point out specifically the increase in superconducting transition temperature of Zr by the ferromagnetic elements in Table I. Considering the empirical fact that superconductivity and ferromagnetism seem to be mutually incompatible, the superconductivity of these alloys suggests that most of the *d* shell holes have been filled.[3] This again is a proof of the formerly stated hypothesis[1] that the superconducting transition temperature is primarily determined by the average electron density and not strongly affected by

TABLE I. Superconducting transition temperatures of Zr alloys with VIIIth column elements.

Fe(~1°)	Co(3.9°)	Ni(1.5°)
Ru(5.7°)	Rh(9°)	Pd(7.5°)
Os(5.2°)	Ir(5.5°)	Pt(3°)

[1] B. T. Matthias, Phys. Rev. **97**, 74 (1955).
[2] J. G. Daunt, Phys. Rev. **80**, 911 (1950).
[3] This is consistent with the fact that Fe, Co, and Ni dissolve substitutionally rather than interstitially in Zr.

the individual atom. (Mass and volume are of course important, but are neglected for the moment to preserve the simplicity of the picture.) Thus, an alloy containing 5% Fe and 5% Ni shows a behavior similar to that of an alloy with 10% Co. In the same way, an alloy with 5% Ru and 5% Pd duplicates the one with 10% Rh.

Carrying this concept further, it was possible to obtain a superconducting compound which seemed comparable to $CoSi_2$, which becomes superconducting at 1.4°K. It is $CoZr_2$ which becomes superconducting at 6.3°K. We found also that $NiZr_2$ becomes superconducting at 1.52°K.

We wish to thank Dr. J. K. Galt for many helpful comments.

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Magnetic Moment Arrangements and Magnetocrystalline Deformations in Antiferromagnetic Compounds*

YIN-YUAN LI
Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received June 27, 1955)

The neutron diffraction data of Shull *et al.* on the powder specimens of MnO, FeO, CoO, etc., are re-examined. It is found that except for FeO the observed Debye-Scherrer patterns may be produced by a number of magnetic moment arrangements with the same antiparallel coupling among the moments on the next nearest neighboring ions, but with different sublattice correlations and different orientations of magnetic axis (or axes). The symmetry change observed at the Néel point is interpreted as a strong effect of spontaneous anisotropy magnetostriction. As a consequence the uniaxial direction of the modified symmetry must coincide with that of the magnetic axis. Assignments of antiferromagnetic arrangements for the compounds are reached. Our prescription does not agree with that of Shull *et al.* except in the case of FeO. The sharp readjustment of lattice parameters at the transition temperature and the orientation of spin axis in other antiferromagnetic compounds are discussed from the viewpoint of the anisotropy magnetostriction approach.

I. INTRODUCTION

IN this note we intend to give a close re-examination of the neutron diffraction data obtained by Shull *et al.*[1] on MnO, FeO, CoO, NiO, MnS, and MnSe in powder form. This study is of interest in view of the fact that whereas FeO, MnO, NiO, and MnS are distorted slightly from the cubic to rhombohedral symmetry and CoO becomes tetragonal at temperatures below the Néel point, yet the arrangements of moments on magnetic ions assigned by Shull *et al.* to MnO, NiO, MnS, and CoO are all the same, both as to the configuration of + and − spins and the direction of spin axis.[2] It should be remarked that the symmetry changes stated above are established results of more than one author[3,4] using x-ray techniques. The explanation of this magnetocrystalline effect put forward by Smart and Greenwald[5] is not consistent with the same discrepancy. From our analysis of the neutron diffraction data and the hypothesis of anisotropy magnetostriction we prescribe for these compounds (except FeO) antiferromagnetic arrangements different from those assigned previously. We shall also discuss the crystallographic position of the magnetic axis in other antiferromagnetics such as Cr_2O_3, MnF_3, etc.

The neutron diffraction patterns of MnO, MnS, MnSe, CoO, and NiO below their respective Néel points are alike. The pattern of FeO is slightly different with the singular absence of a certain magnetic diffraction line. Shull *et al.* find that the magnetic unit cell dimensions are twice as large as the dimensions of a chemical unit cell and that the *next nearest* neighboring magnetic ions must have the spins opposite. These statements are conclusive and can be verified with little effort. However, there are two types of spin arrangements, both consistent with the next nearest antiparallel coupling, but different in the coupling among the *nearest* neighboring magnetic ions. They are illustrated in Fig. 1. In type *A* the four face-centered cubic (f.c.c.) sublattices of antiparallel spins are correlated such that on each (111) plane the spins are parallel and the spins on two neighboring (111) planes opposite; in type *B* this particular coupling among the nearest spins is absent. The four f.c.c. sublattices divide the full f.c.c. lattice in such a way that no two *nearest* neighboring ions are in the same sublattice. Shull *et al.* have shown that a satisfactory agreement is obtained between the

* This research was supported by the Army Signal Corps and the Office of Naval Research under contracts with the Carnegie Institute of Technology.

[1] Shull, Strauser, and Wollan, Phys. Rev. **83**, 333 (1951).

[2] Since we do have in some cases orbital contributions to the magnetic moment, it is preferable to use the terms "antiferromagnetic axis" or "magnetic axis" instead of "spin axis." We shall use them alternatively with the same meaning. Similarly, "spins" and "magnetic moments" are used interchangeably.

[3] H. P. Rooksby, Acta Cryst. **1**, 226 (1948); N. C. Tombs and H. P. Rooksby, Nature **165**, 442 (1950); H. P. Rooksby and N. C. Tombs, Nature **167**, 364 (1951).

[4] S. Greenwald and J. S. Smart, Nature **166**, 523 (1950); S. Greenwald, Acta Cryst. **6**, 396 (1953).

[5] J. S. Smart and S. Greenwald, Phys. Rev. **82**, 113 (1951).

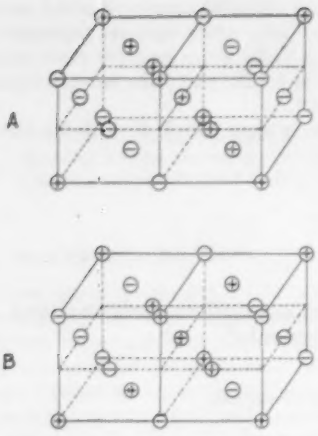


FIG. 1. Two different arrangements of magnetic moments with antiparallel coupling between the *next nearest* neighbors on the f.c.c. lattice.

observed diffraction pattern of MnO and the calculated peak intensities by assuming a configuration of type A with the spin axis in a cubic direction, and that the pattern in FeO can be accounted for by assuming a type A arrangement with a [111] spin axis. They also conclude that CoO, NiO, MnS, and MnSe have the same spin arrangement as MnO. They recognize the alternative arrangement of type B but reject it on the grounds that no distortion of the cubic symmetry would be expected at the antiferromagnetic transition. We shall take a different point of view on this matter in the coming discussion.

II. DEBYE-SCHERRER INTENSITY AND ALTERNATIVE ASSIGNMENTS OF MAGNETIC STRUCTURE

Before we proceed to state our viewpoint, it is necessary to consider, in detail, the Debye-Scherrer intensity of these compounds. Since unpolarized neutrons were used in the experiments, the observed intensity is the sum of nuclear and magnetic diffractions. The former is independent of the magnetic structure and so we need only to discuss the magnetic diffraction intensity factor $|\mathbf{q}F_m|^2$, where F_m is the magnetic structure amplitude of a magnetic unit cell and $\mathbf{q}$ is $\boldsymbol{\kappa}(\boldsymbol{\kappa}\cdot\mathbf{n})-\mathbf{n}$, where $\boldsymbol{\kappa}$ and $\mathbf{n}$ are respectively the unit vector in the direction of the scattering vector and the magnetic axis. We have for the type A configuration

$$F_m=8f_m\{1-\exp i\pi(h+k)/2-\exp i\pi(k+l)/2-\exp i\pi(l+h)/2\}, \quad (1A)$$

and for type B

$$F_m=8f_m\{1+\exp i\pi(h+k)/2-\exp i\pi(k+l)/2-\exp i\pi(l+h)/2\}, \quad (1B)$$

where

$$f_m=(e^2/mc^2)\nu Sf \quad (2)$$

is the magnetic scattering amplitude of the magnetic ion concerned, whose spin quantum number is S. ν is the magnetic moment of the neutron, and f is an atomic form factor which takes care of the decrease of the scattered amplitude as a result from the finite extension of the magnetically active electrons. h, k, and l are the indices referred to the magnetic unit cell and are all odd numbers in the equations. ($F_m=0$ for h, k, and l not all odd.) The factor 8 is found in the equations, because for odd indices the diffractions by the 8 chemical unit cells in a magnetic unit cell are coherent. Let α, β, and γ be the direction cosines of the magnetic axis. We write

$$q^2=1-(\alpha h+\beta k+\gamma l)^2/N, \quad (3)$$

where

$$N=h^2+k^2+l^2.$$

For type A we derive:

$$|F_m/8f_m|^2=\begin{cases}16, \text{ when } \left.\begin{matrix}h-k\\k-l\\l-h\end{matrix}\right\}\text{are divisible by 4,}\\0, \text{ otherwise;}\end{cases} \quad (4A)$$

and

$$q_m^2=\tfrac{2}{3}\{1-(\alpha\beta+\beta\gamma+\gamma\alpha)(hk+kl+lh)/N\}. \quad (5A)$$

For type B:

$$|F_m/8f_m|^2=4, \quad (4B)$$

and

$$q_m^2=\tfrac{2}{3}, \quad (5B)$$

independent of α, β, and γ. q_m^2 is the average of q^2 calculated over all the possible permutations and combinations of signs of $(\pm h,\pm k,\pm l)$ with given h, k, and l for which F_m is not zero. For type A three out of four of all these permutations and combinations of signs do not satisfy the condition that $h-k$, $k-l$, and $l-h$ be divisible by 4. They make no contribution to the intensity. Those with $h-k$, $k-l$, and $l-h$ divisible by 4 have an intensity four times that of each diffraction maximum of $B[\alpha\beta\gamma]$. Therefore, the spin arrangements $A[100]$ and $B[\alpha\beta\gamma]$ must yield the same Debye-Scherrer pattern, although their diffraction patterns produced by a single crystal under neutron irradiation should be markedly different.

The occurrence of more than one magnetic axis among the sublattices cannot be ruled out by physical considerations. Its influence on the susceptibilities has been discussed by Van Vleck.[6] We have evidence to believe that this is the case in MnO_2.[7] Therefore we must, in general, allow that the magnetic axes of the four sublattices in the MnO type compounds may be all in different orientations. Let $\mathbf{n}_i(\alpha_i,\beta_i,\gamma_i)$ with $i=0,1,2,3$ be the unit vectors along the four magnetic axes and

[6] J. H. Van Vleck, J. phys. radium **12**, 262 (1951).
[7] R. A. Erickson, Phys. Rev. **85**, 745 (1952); H. Bizette and B. Tsai, *Colloque sur les Phenomenes Cryomagnetiques* (Perrin-Langevin, Paris, 1948).

$\mathbf{q}_i=\boldsymbol{\kappa}(\boldsymbol{\kappa}\cdot\mathbf{n}_i)-\mathbf{n}_i$. Now we have

$$|\sum_i \mathbf{q}_i(F_m)_i|^2=(8f_m)^2|\,\mathbf{q}_0+\mathbf{q}_1\exp i\pi(k+l)/2 +\mathbf{q}_2\exp i\pi(l+h)/2+\mathbf{q}_3\exp i\pi(h+k)/2|^2. \quad (6)$$

We derive the intensity of a Debye-Scherrer line

$$I(hkl)\propto\sum_{(\pm h,\pm k,\pm l)}|\textstyle\sum_i\mathbf{q}_i(F_m)_i|^2=4(8f_m)^2pQ_{hkl}^2, \quad (7)$$

where $\sum_{(\pm h,\pm k,\pm l)}$ is summed over all the possible permutations and combinations of signs of $(\pm h,\pm k,\pm l)$, and p is the multiplicity factor.

$$Q_{hkl}^2=\tfrac{2}{3}[1-\phi(\mathbf{n}_0,\mathbf{n}_1,\mathbf{n}_2,\mathbf{n}_3)(h^*k^*+k^*l^*+l^*h^*)/N], \quad (8)$$

with

$$\phi(\mathbf{n}_0,\mathbf{n}_1,\mathbf{n}_2,\mathbf{n}_3)=\tfrac{1}{4}(\alpha_1\beta_2+\beta_1\alpha_2+\beta_2\gamma_3+\gamma_2\beta_3+\gamma_3\alpha_1+\gamma_1\alpha_3 -\alpha_0\beta_3-\beta_3\alpha_2-\beta_1\gamma_0-\gamma_1\beta_0-\gamma_3\alpha_0-\alpha_2\gamma_0), \quad (9)$$

and (h^*,k^*,l^*) is one of those combinations of $(\pm h,\pm k,\pm l)$ such that h^*-k^*, k^*-l^*, and l^*-h^* are all divisible by 4. When we put $\mathbf{n}_0=\mathbf{n}_1=-\mathbf{n}_2=-\mathbf{n}_3$ the general case is reduced to the type B arrangement discussed above, and the right-hand side of Eq. (9) is zero. Similarly, when $\mathbf{n}_0=-\mathbf{n}_1=-\mathbf{n}_2=-\mathbf{n}_3=\mathbf{n}(\alpha,\beta,\gamma)$ we have the type A arrangement and $\phi=(\alpha\beta+\beta\gamma+\gamma\alpha)$. In both cases Q_{hkl}^2 reduces to q_{hkl}^2 given by Eq. (4A) and (4B) respectively. In general any arrangement making $\phi=0$ must produce the same Debye pattern as that of $A[100]$ and $B[\alpha\beta\gamma]$ and those satisfying $\phi=1$ must produce a powder diffraction pattern, the same as that of $A[111]$. Among the former we are specially interested in the case of $\mathbf{n}_0=\mathbf{n}_1=(100)$ and $\mathbf{n}_2=\mathbf{n}_3=(010)$. For this arrangement the notation $C[100][010]$ will be used.

It is inevitable to conclude that with the powder diffraction data alone a large number of spin arrangements may be assigned to each of the MnO type compounds. When we assume a single magnetic axis for the four sublattices, there is no ambiguity in the antiferromagnetic structure of FeO which must be $A[111]$, but even then the spin arrangement of MnO, NiO, CoO, MnS, and MnSe still cannot be decided. The analysis given in the preceding paragraph thoroughly illustrates the limitation of the Debye-Scherrer method in neutron diffraction work, a situation similar to that of the same method in the x-ray analysis. A precise assignment of a magnetic structure cannot be reached with the results of the powder diffraction technique alone, when there are more than two sublattices of parallel spins involved and the direction of spin axis (or axes) is not revealed by the singular absence of a certain diffraction line. In our case there are eight sublattices of parallel spins. In FeO we notice the singular absence of the (111) line.

We shall now look into the symmetry changes accompanying the magnetic transition. The rhombohedral distortion of the cubic structure of MnO and NiO produces a contraction in the [111] direction and that of the FeO an expansion in the same direction. In the type A arrangement a contraction or an expansion in the trigonal direction changes the distance between the antiferromagnetically coupled spins on the neighboring (111) planes. Smart and Greenwald have proposed to explain the symmetry distortion by assuming that the dependence of the exchange integral J on the distance between magnetic ions is in favor of such a contraction or an expansion. However, the tetragonal deformation of CoO directly contradicts this exchange magnetostriction hypothesis. We shall adopt the viewpoint that this magnetocrystalline effect is simply a strong manifestation of (spontaneous) anisotropy magnetostriction. The deformation amounts to a strain of 10^{-3} to 10^{-2} which is larger than the usual magnetostriction by several orders of magnitude. The term anisotropy magnetostriction is used to describe the fact that the anisotropy energy depends on the strain, and, as a consequence of this and the minimization of the internal energy, the lattice of the antiferromagnetic crystal must be a deformed one with respect to the cubic lattice of the paramagnetic state.[8] In this mechanism it is the direction of magnetic axis rather than the configuration in which the moments on magnetic ions are coupled that determines the characteristics of the magnetocrystalline deformation. By symmetry considerations we expect that the compounds with rhombohedral symmetry, when antiferromagnetic, are the ones having the spin axis in the [111] direction. MnO, MnS, and NiO must have the type B arrangement but FeO the type A correlation of sublattices. Similarly, we may assign to CoO either the A or $B[001]$ arrangement or $C[100][010]$, in each case [001] being the tetragonal axis. For the purpose of clarity we list in Table I the compounds, their symmetry below the Néel point, and the two antiferromagnetic structures prescribed respectively by Shull *et al.* and by the present writer. Because of the large number of possibilities that lead to a given value of ϕ of Eq. (9), our assignments given in Table I may not be the only ones consistent with the neutron diffraction data and the anisotropy magnetostriction hypothesis. It is interesting to note that according to our assignment the $B[111]$ structure is distorted by a contraction in the trigonal direction ($\alpha>60°$) and the $A[111]$ structure by an expansion in the same direction

TABLE I. Crystalline deformation and assignment of magnetic moment arrangement in the MnO type compounds.

Compounds	MnO	FeO	CoO	NiO	MnS	MnSe
Symmetry[a]	Rh.	Rh.	Tet.	Rh.	Rh.	Rh.?
Rhombohedral angle, or tetragonality	$\alpha>60°$	$\alpha<60°$	$c/a<1$	$\alpha>60°$	?	?
Shull *et al.*[b]	$A[001]$	$A[1\bar{1}1]$	$A[001]$	$A[001]$	$A[001]$	$A[001]$
This note	$B[111]$	$A[111]$	$B[001]$ or $A[001]$ or $C[100][010]$	$B[111]$	$B[111]$	$B[111]$

[a] Rh. = rhombohedral; Tet. = tetragonal.
[b] See reference 1.

[8] R. Becker and W. Döring, *Ferromagnetismus* (Verlag Julius Springer, Berlin, 1939), Chap. III, Sec. 11.

($\alpha<60°$). A possible explanation of the different sign of $\Delta\alpha$ for FeO is that in the $A[111]$ structure exchange magnetostriction must also be expected to occur but not in the $B[111]$ structure.

The issue between the different assignments can be decided either by neutron diffraction on a single crystal or by measuring the susceptibilities of a single crystal to determine the direction of spin axis. Indirect evidence concerning the spin arrangement can be obtained by conducting the following suggested experiments which may be performed either on a sintered specimen or single crystal. The sharp decrease of Young's modulus of CoO and NiO at the Néel point as shown by Street and Lewis[9] and of CoO by Fine[10] indicates the readiness of the trigonal axis in NiO and the tetragonal axis in CoO to switch to a more favorable orientation, presumably a crystallographically equivalent one, under an applied stress. Single crystals of NiO are always badly twinned even under extremely careful handling and the distribution of twinned domains can be disturbed by a light touch.[11] In Rh. $A[100]$ structure the spin axis is not at all related to the rhombohedral direction, while in Rh. $A[111]$, Rh. $B[111]$, and Tet. A or $B[001]$ the spin axis coincides with the rhombohedral or tetragonal axis. Therefore, in the former the spin axis would not have to turn when the rhombohedral axis flops under external stress, but in the latter the spin axis must turn with the uniaxis of symmetry. The change of the direction of spin axis can be detected by the difference in susceptibility before and after the application of external stress. On the other hand, with a sufficiently strong magnetic field we could switch the spin axis from one of its normal direction to an equivalent one. In Rh. $A[111]$, Rh. $B[111]$, and Tet. $B[001]$ the uniaxis of the crystal must turn with the spin axis and produce an elastic strain. In Rh. $A[100]$ the three cubic directions are equivalent with respect to the rhombohedral axis. The switch of spin axis would not demand a flop of the rhombohedral axis, say, from $[111]$ to $[\bar{1}11]$ and so no strain would be induced. According to the following discussion an extremely strong field is required in this experiment.

III. DISCUSSION

Our anisotropy magnetostriction hypothesis must face the question of why the deformation is such a large effect in the MnO type compounds as well as other antiferromagnetics to be mentioned later, being larger than the magnetostriction in the ordinary ferromagnetics by orders of magnitude. We may answer this by affirming that the large magnetostriction is tied to a high anisotropy in the antiferromagnetics, since both are under the same influence of a strong spin-orbit coupling. The argument goes as follows:

The dipole-dipole term of the anisotropy energy has been calculated by Kaplan[12] for type A arrangement on an undeformed cubic lattice. He obtains

$$E_d = K_d(\alpha\beta+\beta\gamma+\gamma\alpha), \tag{10A}$$

where

$$K_d = 1.64\ 10^7 \text{ ergs/cm}^3 \text{ for MnO}. \tag{11}$$

For type B arrangement our calculation shows

$$E_d = 0 \tag{10B}$$

independent of the direction of spin axis. Evidently E_d of Eq. (10) in either case cannot be taken as the anisotropy energy we wish to discuss, since we have not considered the deformed symmetry, nor included the second order effect of the dipole interaction, nor the multipole interactions.[13] Moreover, we must also take into consideration the spin-orbit coupling. A calculation of the anisotropy energy including these effects would be enormously troublesome, even to obtain a quite approximate result. Fortunately, the following observation can be made from Eq. (10A). For the spin arrangement of FeO $A[111]$, E_d is a maximum of Eq. (10A) instead of a minimum. Therefore, this dipole-dipole energy is not the dominant term of the total anisotropy, although this term already amounts to 10^7 erg/cm^3. Accordingly, we must conclude that for the MnO type compounds the anisotropy energy is of the order of 10^8–10^9 erg/cm^3. (Incidentally, we may mention that in order to overcome such a high anisotropy a field of 10^5 gauss or higher would be required to turn the spin axis.) The origin of this high anisotropy should be sought in the Mn-O-Mn super-exchange. The importance of spin-orbit coupling in the calculation of anisotropy energy has been shown by Van Vleck.[13] When this coupling is not taken into consideration the second-order perturbation calculation of the dipole energy of a cubic ferromagnetic gives a result only 1/1000 of the observed anisotropy. The anomalous magnitude of the anisotropy can be reasonably attributed to the effect of the spin-orbit coupling with the anisotropy in the energy of orbital valence. This applies as well whether or not the inclusion of the multipole interactions besides the dipole term is necessary. Vonsovsky[14] has shown that with the orbital wave function acting as the bridge between the spin and the crystal lattice, the dependence of the dipole energy and the spin-orbit energy can account for the observed magnetostriction in the metals of the iron group. Presumably, a strong spin-orbit coupling should lead to a large magnetostriction. On the grounds of these connections we may hold that a large (spontaneous) magnetostriction should be found in magnetics of high anisotropy such as the MnO-type antiferromagnetics.

We shall present the following empirical findings to

[9] R. Street and B. Lewis, Nature **168**, 1036 (1951).
[10] M. E. Fine, Revs. Modern Phys. **25**, 158 (1953).
[11] W. Roth (unpublished).
[12] J. I. Kaplan, J. Chem. Phys. **22**, 1709 (1954).
[13] J. H. Van Vleck, Ann. inst. Henri Poncaré **10**, 57 (1947).
[14] S. V. Vonsovsky, J. Phys. (U.S.S.R.) **3**, 181 (1940).

demonstrate some remarkable correlations between the magnetostriction and the anisotropy energy. The magnetostrictive energy (the sum of the elastic strain energy and the magnetoelastic energy) may be expressed by an increment ΔK in the anisotropy constant. ΔK is a function of the elastic moduli and the magnetostrictive constants of the crystal. This has been explicitly shown by Kittel[15] for ferromagnets of cubic symmetry. For ferromagnetic metals and alloys the ratio $\Delta K/K$ is generally of the order of 10^{-1} to 10^{-3} and K is of the order of 10^5 erg/cm^3. For the MnO type compounds $K \approx 10^8 - 10^9$ erg/cm^3. ΔK is of the order of $Y\lambda^2$, where Y is the Young's modulus and λ the magnetostrictive strain. For CoO Greenwald[4] found $\lambda \approx 10^{-2}$ to 10^{-3} increasing with the decrease of temperature, and according to Street and Lewis[9] $Y \approx 10^{12}$ dyne/cm^2. Therefore, here again we have $\Delta K/K \approx 10^{-1}$ to 10^{-3}. It seems that the anisotropy energy and the energy of magnetostriction varies roughly in proportion from metallic ferromagnetics to the antiferromagnetic compounds.

In order to test our confidence in the anisotropy magnetostriction approach we shall examine the relation between the magnetic axis (or axes) and the crystal summetry of the antiferromagnetics for which such information is available.[16] With the exception of $CuCl_2 \cdot 2H_2O$, which is in a different category, they are listed in Table II. No change of symmetry has been reported at the Néel point of these compounds. For several cases sharp changes of the lattice parameters[17] are found accompanying the magnetic transition. Probably all the compounds have, more or less, a readjustment of lattice spaces at the Néel point. Except the state of α-Fe_2O_3 in the temperature range of $-20°C$ and 675°C, their spin axis is in the uniaxial direction, or in the case of MnO_2 the spin axes are orientated symmetrically in the plane perpendicular to the uniaxial direction. This situation agrees perfectly with the concept of anisotropy magnetostriction. In the case of NiF_2, the deviation in the direction of spin axis from the tetragonal axis by 10 degrees should induce a slight distortion of the true tetragonal symmetry. An accurate x-ray analysis should be carried out on this compound.

TABLE II. Crystalline structure and the orientation of magnetic axis (or axes) in antiferromagnetic compounds.

Compounds	Crystal structure	Directions of spin axis (or axes)
MnF_2, FeF_2, CoF_2, and NiF_2[a]	Tetragonal, rutile structure	Tetragonal axis
MnO_2	Tetragonal, rutile structure	Two spin axes within the xy-plane and mutually perpendicular
CrSb	Hexagonal, AsNi structure	Hexagonal axis
α-Fe_2O_3 (below $-20°C$) and Cr_2O_3	Rhombohedral α-Al_2O_3 structure	Rhombohedral axis
α-Fe_2O_3 (between $-20°C$ and 675°C)	Rhombohedral α-Al_2O_3 structure	Within the (111) plane and direction to a nearest neighboring Fe^{+3} in the plane

[a] In NiF_2 the spin alignment is not exactly along the tetragonal axis but makes an angle of 10 degrees with the latter.

The degree of distortion depends, of course, on the strength of anisotropy-elastic coupling, which is not known for this compound. The spin axis in the high-temperature antiferromagnetic state of α-Fe_2O_3 is perpendicular to the trigonal axis. From our viewpoint the crystal structure should become monoclinic in the temperature range of this state. Yet, according to the x-ray analysis of Willis and Rooksby,[17] α-Fe_2O_3 remains rhombohedral with a sharp change of lattice parameters at 675°C. This should not be taken as a negative evidence of the general validity of our hypothesis, as can be explained in the following. Anderson *et al.*[18] find by a resonance technique that the anisotropy force in α-Fe_2O_3 is only about 60 gauss in the (111) plane, while that in a plane containing the [111] axis is 30 000 gauss. Combining this experimental fact and the concept mentioned in the preceding paragraph that a strong magnetostriction goes together with high anisotropy, we must anticipate in α-Fe_2O_3 a small magnetostriction *within* the (111) plane. Therefore, an x-ray experiment cannot detect the distortion of the (111) cross section of a unit cell, although the readjustment of lattice parameters is so much larger and more readily measured by x-ray technique. Another effect of concurrent interest is the $-20°C$ transition of α-Fe_2O_3 at which the magnetic axis turns to the [111] direction in the lower temperature range. Willis and Rooksby report that the change, if any, of lattice parameters at this transition is considerably less than that which occurs at the high Néel temperature associated with the magnetic lattice formation. This simply means that the two antiferromagnetic states of α-Fe_2O_3, which are different only in the orientation of magnetic axis, have approximately the same spontaneous magnetostriction. No criterion on our anisotropy magnetostriction hypothesis can be derived from this effect.

The author wishes to thank Professor J. E. Goldman for his kind interest in this work. He is also indebted to Dr. J. S. Smart and Dr. C. G. Shull for helpful comments.

[15] C. Kittel, Revs. Modern Phys. **21**, 541 (1949).

[16] The sources of information for the antiferromagnetics listed are: (a) α-Fe_2O_3: see reference 1. (b) Cr_2O_3: B. N. Brockhouse, J. Chem. Phys. **21**, 961 (1953); McGuire, Scott, and Grannis, Phys. Rev. **98**, 1562(A) (1955). (c) MnF_2: R. A. Erickson, Phys. Rev. **90**, 779 (1953). (d) MnO_2: See reference 7. (e) CrSb: A. I. Snow, Phys. Rev. **85**, 365 (1953); Revs. Modern Phys. **25**, 127 (1953).

[17] The sources of information on the change of lattice parameters are: (a) α-Fe_2O_3: B. T. M. Willis and H. P. Rooksby, Proc. Phys. Soc. (London) **B65**, 950 (1952). (b) Cr_2O_3: S. Greenwald, Nature **168**, 379 (1951). In this article Greenwald reported a contraction in the [111] direction when Cr_2O_3 is cooled below the Néel point. It has been found by R. J. Davis and W. E. Armstrong, and by Greenwald herself (reindexing her original data) that the readjustment of lattice parameters involves an expansion in the [111] direction instead of a contraction. See S. Greenwald, Nature (to be published). (c) CrSb: See reference 16(e), or B. T. M. Willis, Acta Cryst. **6**, 425 (1953).

[18] P. W. Anderson *et al.*, Phys. Rev. **93**, 717 (1954).

Faraday Effect in Germanium at Room Temperature*†

RICHARD R. RAU‡ AND M. E. CASPARI
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania

(Received June 23, 1955)

When a plane polarized electromagnetic wave passes through a semiconductor and a static magnetic field is applied along the direction of propagation, there occurs a rotation of the plane of polarization and the transmitted radiation is found to be elliptically polarized. This effect is due to the influence of the free charge carriers in the semiconductor and has been analyzed using the Drude-Zener model. For small losses, weak magnetic fields, and small values of $\omega\tau$ (assuming the relaxation time τ to be energy independent) the angle of rotation of the plane of polarization can be expressed to a first order of approximation by the simple formula:

$$\theta = \tfrac{1}{2}(\mu_0/\epsilon_0)^{\frac{1}{2}}(\sigma_0\mu B/\sqrt{K'})t,$$

where μ is the Hall mobility, σ_0 is the dc conductivity, B is the magnetic field, t is the thickness of sample traversed, K' is the dielectric constant of the material at the frequency employed in the experiments, and ϵ_0 and μ_0 are the dielectric constant and the permeability of free space respectively. For spherical energy surfaces, the degree of ellipticity, which is a second-order effect, can be expressed by the relation $\mathcal{E} = (\mu_0/\epsilon_0)^{\frac{1}{2}}[\sigma_0(\mu B)(\omega\tau)/\sqrt{K'}]^2$ where $\omega/2\pi$ is the frequency and τ is the relaxation time. Thus, for small losses the ellipticity is proportional to the latter quantity. For the case of low frequencies, the effect can be explained by the introduction of a Hall-effect type field into Maxwell's equations. In general, the angle of rotation and the ellipticity may depend on the direction of the applied magnetic field because of the nonspherical nature of the energy surfaces.

Room-temperature measurements of the angle of rotation at microwave frequencies on both *n*- and *p*-type samples of germanium gave values of 3780 cm²/volt-sec and 3300 cm²/volt-sec for the electron and hole mobilities respectively. The method should be applicable to the determination of mobilities in powdered samples without using electrodes, if the field inside the powdered particles is determined by a Clausius-Mosotti type approximation.

I. INTRODUCTION

THE rotation of the plane of polarization of a plane polarized electromagnetic wave of optical frequencies passing through a substance under the influence of a static magnetic field along the direction of propagation is known as the Faraday effect and has been extensively investigated, both experimentally and theoretically.[1] The effect can be explained by considering the influence of the magnetic field on the equations of motion of the bound electrons which gives rise to different velocities of propagation for the right- and left-handed circularly polarized components of the plane polarized wave. The effect has also been observed in ferromagnetic materials at microwave frequencies.[2] In this case the mechanism is the influence of the precession of the electron spin. Ferromagnetic materials with low loss (ferrites) have found use in microwave circuits.[3]

When a plane-polarized electromagnetic wave passes through a semiconductor and a static magnetic field is applied along the direction of propagation, the plane of polarization is found to rotate and the transmitted radiation becomes elliptically polarized. In this case the effect is caused by the free charge carriers and the mechanism is similar to that producing the Hall effect. Faraday rotation in artificial dielectrics has been predicted by Wicher[4] from an analysis in terms of the Hall field. Using large magnetic fields (~16 kilo-oersted) the Faraday rotation in germanium has been investigated at microwave frequencies by Suhl and Pearson.[5] Their work was done at low temperatures (77°K) where the collision frequency of the free charge carriers is small compared to the signal frequency. Estimates of the effective masses of these carriers could be computed from their measurements.

The present work is concerned with the investigation of the Faraday effect in semiconductors at room temperature (where the collision frequency is greater than the signal frequency) with relatively small magnetic fields, in order to establish what information can be gained from such a simple experiment. The angle of rotation of the plane of polarization of microwaves traversing thin (*ca* 4 mm) samples of *n*- and *p*-type germanium inserted in a circular guide (TE_{11} mode) has been measured. For magnetic fields of 1400 gauss the angle of rotation was about 3 degrees. A simple analysis in terms of a dielectric medium having a concentration of free charge carriers characterized by a mobility and an isotropic effective mass (Drude-Zener model) shows, that the angle of rotation depends on the dc conductivity and the mobility of the charge carriers, as well as on the static dielectric constant, the magnetic field and the sample thickness. Using known values for the static dielectric constant and the dc conductivity, the mobility of the charge carriers can be calculated from our experiments. The relaxation time or the effective mass of the charge carriers does not affect the angle of rotation to a first order of approximation. A knowledge of an accurate value of the relaxation time is, therefore, unnecessary in order to calculate the mobility to great accuracy. If the approximation of

* This work was supported by the U. S. Air Force.

† This paper represents part of a thesis submitted by one of us (R.R.R.) to the Department of Physics, University of Pennsylvania, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

‡ Now in the U. S. Army stationed in Greenland.

[1] See for instance: W. Schütz, *Handbuch der Experimentalphysik* (Akademische Verlagsgeschellschaft MBH, Leipzig, 1936), pp. 1–199.

[2] C. Hogan, Revs. Modern Phys. **25**, 253 (1953).

[3] C. Hogan, Bell System Tech. J. **31**, 1 (1952).

[4] E. Wicher, J. Appl. Phys. **22**, 1327 (1951).

[5] H. Suhl and G. L. Pearson, Phys. Rev. **92**, 858 (1953).

low losses is made, the above analysis predicts a degree of ellipticity of the transmitted radiation which depends linearly on the relaxation time. Both effects are complicated by the anisotropy of the effective masses of the charge carriers, especially when the applied magnetic field is large.

II. ANALYSIS OF THE MECHANISM BY THE DRUDE-ZENER THEORY[6,7]

The Faraday effect in semiconductors can be readily analyzed in terms of the classical Drude-Zener model. The equations of motion of the electrons or holes under the influence of the static magnetic field and the microwave electromagnetic field are identical in this analysis to those of importance in the classical theory of cyclotron resonance in semiconductors, as developed, for instance, in the publications of Dresselhaus, Kip, and Kittel[8] and Lax, Zeiger, and Dexter.[9]

The effect will first be analyzed for a model of simple spherical energy surfaces. Consider the propagation of plane electromagnetic waves in a medium of dielectric constant ϵ_{sl}'(real) containing N free charge carriers per unit volume, each with a charge q, a single effective mass m, and a relaxation time τ[10] subjected to a static magnetic field B_z in the direction of propagation. The charge carriers obey the following equation of motion:

$$m\frac{d\mathbf{v}}{dt}+\frac{m}{\tau}\mathbf{v}=q(\mathbf{E}+\mathbf{v}\times\mathbf{B}), \tag{1}$$

where $\mathbf{v}$ is the drift velocity of the charge carriers.

If the motion of the charge carriers, as well as that of the electromagnetic field is confined to the (xy) plane (the static magnetic field being applied in the z direction) we have, assuming isotropic scattering:

$$m\dot{v}_{\pm}+\frac{m}{\tau}v_{\pm}=qE_{\pm}\mp iqB_z, \tag{2}$$

where

$$\begin{aligned} v_{\pm}&=v_x\pm iv_y,\\ E_{\pm}&=E_x\pm iE_y. \end{aligned} \tag{3}$$

For steady-state solutions under harmonic electric fields the total current density including contributions from these charge carriers and the displacement current is

$$J_{\text{total}\pm}=\left[\epsilon_{sl}'-\frac{Nq^2/\omega(m\omega\pm qB_z)}{(m/\tau)^2+(m\omega\pm qB_z)^2}\right]\dot{E}_{\pm}+\frac{(Nq^2m/\tau)E_{\pm}}{(m/\tau)^2+(m\omega\pm qB_z)^2}, \tag{4}$$

where ω is the angular frequency of the impressed radiation. Thus, corresponding to the two values E_+ and E_- for the electric field, there are two values for the effective dielectric constant and conductivity respectively:

$$\begin{aligned} \epsilon_{\pm}'&=\epsilon_{sl}'-\frac{\sigma_0/\omega(\omega\tau\pm\mu B_z)}{1+(\omega\tau\pm\mu B_z)^2},\\ \sigma_{\pm}&=\frac{\sigma_0}{1+(\omega\tau\pm\mu B_z)^2}, \end{aligned} \tag{5}$$

where $\sigma_0=Nq^2\tau/m$ is the dc conductivity and $\mu=q\tau/m$ is the mobility. The fields E_+ and E_- are solutions of Maxwell's equations in a nonmagnetic medium and for plane wave solutions propagating in the $+z$ direction, we have

$$k_{\pm}^2=\mu_0(\epsilon_{\pm}'\omega^2-\sigma_{\pm}\omega). \tag{6}$$

Here μ_0 is the permeability of free space and $k_{\pm}=\alpha_{\pm}+i\beta_{\pm}$ are the complex propagation constants corresponding to right- and left-handed circularly polarized light respectively. The difference between these propagation constants results in the production of elliptically polarized waves from plane polarized waves with the major axis of the ellipse rotated with respect to the original plane of polarization. It can be seen that for either a reversal of the direction of the magnetic field or a reversal of the sign of the carriers, the two propagation vectors are interchanged.

The angle of rotation θ and the degree of ellipticity δ of the transmitted radiation can be expressed in terms of the real and imaginary parts of the propagation constants.[11] For small thicknesses t,

$$\theta=\tfrac{1}{2}(\alpha_--\alpha_+)t, \tag{7}$$

and

$$\delta=\tfrac{1}{2}(\beta_+-\beta_-)t. \tag{8}$$

Hence, the angle of rotation and the degree of ellipticity are very nearly proportional to the sample thickness t (for small t) and the phenomenon can, therefore, be described as a type of "Faraday effect."

It is shown in treatises on electromagnetic theory[12] that

$$\alpha_{\pm}=\omega(\epsilon_0\mu_0)^{\frac{1}{2}}\left\{\frac{1}{2}\frac{\epsilon_{\pm}'}{\epsilon_0}\left[\left(1+\frac{\epsilon_{\pm}''^2}{\epsilon_{\pm}'^2}\right)^{\frac{1}{2}}+1\right]\right\}^{\frac{1}{2}} \tag{9}$$

and

$$-\beta_{\pm}=\omega(\epsilon_0\mu_0)^{\frac{1}{2}}\left\{\frac{1}{2}\frac{\epsilon_{\pm}'}{\epsilon_0}\left[\left(1+\frac{\epsilon_{\pm}''^2}{\epsilon_{\pm}'^2}\right)^{\frac{1}{2}}-1\right]\right\}^{\frac{1}{2}}, \tag{10}$$

where $\epsilon_{\pm}''=\sigma_{\pm}/\omega$.

From these two equations the angle of rotation and the ellipticity can be calculated using Eqs. (7) and (8) and values of $\epsilon_{\pm}'$ and $\epsilon_{\pm}''$ from Eq. (5). Although the mobility and the relaxation time appear explicitly in

[6] Some the the equations derived in this section are in essence identical to those obtained independently by Suhl and Pearson.[7] The authors are very grateful to Dr. Pearson for submitting to them a copy of his paper delivered before the meeting of the American Physical Society at Rochester, New York, June, 1953.

[7] H. Suhl and G. L. Pearson (private communication).

[8] Dresselhaus, Kip, and Kittel, Phys. Rev. **98**, 368 (1955).

[9] Lax, Zeiger, and Dexter, Physica **20**, 818 (1954).

[10] T. S. Benedict and W. Shockley, Phys. Rev. **89**, 1152 (1953); Phys. Rev. **91**, 1565 (1953).

[11] See for instance: K. Försterling, *Lehrbuch der Optik* (S. Hirzel, Leipzig, 1928), p. 44.

[12] See for instance: A. von Hippel, *Dielectrics and Waves* (John Wiley and Sons, Inc., New York, 1954), p. 28.

Eq. (5) it is possible to show that to a first approximation, the angle of rotation is independent of the relaxation time and hence a value for the mobility can be obtained from a measurement of the angle of rotation without an accurate knowledge of the relaxation time.

As a first approximation, we assume that the conductivity is independent of the frequency so that

$$\epsilon_{\pm}''=\sigma_0/\omega.$$

Expressing the dielectric constant in the form

$$\epsilon_{\pm}'=\epsilon_{zz}'-(\Delta\epsilon)_{\pm},$$

where

$$(\Delta\epsilon)_{\pm}=\frac{\sigma_0/\omega(\omega\tau\pm\mu B_z)}{1+(\omega\tau\pm\mu B_z)^2},$$

we can write Eq. (9) as

$$\alpha_{\pm}^2=\frac{\omega^2\mu_0}{2}\left\{(\epsilon_{zz}'-(\Delta\epsilon)_{\pm})\left[1+\left(\frac{\sigma_0/\omega}{\epsilon_{zz}'-(\Delta\epsilon)_{\pm}}\right)^2\right]^{\frac{1}{2}}+\epsilon_{zz}'-(\Delta\epsilon)_{\pm}\right\}. \quad (11)$$

If we neglect terms involving $(\Delta\epsilon)_{\pm}^2$, obvious reductions give

$$\alpha_-^2-\alpha_+^2=\frac{\omega^2\mu_0}{2}\chi[(\Delta\epsilon)_+-(\Delta\epsilon)_-], \quad (12)$$

where

$$\chi=\chi\left(\frac{\sigma_0}{\omega},\epsilon_{zz}'\right)=\left[1+\left(\frac{\sigma_0/\omega}{\epsilon_{zz}'}\right)^2\right]^{\frac{1}{2}}+1-\frac{1}{1+\left(\frac{\omega\epsilon_{zz}'}{\sigma_0}\right)^2}.$$

Now

$$\alpha_--\alpha_+\cong(\alpha_-^2-\alpha_+^2)/2\alpha_{\text{lw}} \quad (13)$$

with

$$\alpha_{\text{lw}}=\omega(\tfrac{1}{2}\epsilon_0\mu_0)^{\frac{1}{2}}[K_{zz}'\eta-(\Delta\epsilon)_{\text{lw}}\chi]^{\frac{1}{2}},$$

where

$$K_{zz}'=\epsilon_{zz}'/\epsilon_0,\quad (\Delta\epsilon)_{\text{lw}}=\sigma\tau/\epsilon_0,$$

and

$$\eta=\left[1+\left(\frac{\sigma_0/\omega}{\epsilon_{zz}'}\right)^2\right]^{\frac{1}{2}}+1.$$

For weak magnetic fields and small values of $\omega\tau$ (μB_z and $\omega\tau\ll1$), we obtain for the angle of rotation, using Eqs. (5), (7), (12), and (13)

$$\theta=\frac{\chi}{2\sqrt{2}}\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\frac{\sigma_0\mu B_z l}{[\eta K_{zz}'-\chi(\Delta\epsilon)_{\text{lw}}]^{\frac{1}{2}}}. \quad (14)$$

For most semiconductors, even at microwave frequencies, $(\Delta\epsilon)_{\text{lw}}$ can be neglected and we then obtain

$$\theta=\frac{\chi}{2\sqrt{2}}\left(\frac{\mu_0}{\epsilon_0\eta}\right)^{\frac{1}{2}}\frac{\sigma_0\mu B_z l}{\sqrt{K_{zz}'}}. \quad (15)$$

At microwave frequencies for most semiconductors $\sigma_0/\omega\epsilon_{zz}'\simeq1$ so that, as a rough estimate

$$\theta\simeq0.44\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\frac{\sigma_0\mu B_z l}{\sqrt{K_{zz}'}}. \quad (16)$$

For the case of small losses ($\sigma_0/\omega\ll\epsilon'$),

$$\theta=\frac{1}{2}\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\frac{\sigma_0\mu B_z l}{\sqrt{K'}}, \quad (17)$$

where K' is the dielectric constant of the material at the frequency employed. In general, values of sufficient accuracy for the angle of rotation can be obtained from Eq. (15). If greater accuracy is desired a rough estimate of the relaxation time τ (i.e., $(\Delta\epsilon)_{\text{lw}}$) is sufficient to obtain very accurate values for the mobility from measurements of the angle of rotation θ, using Eq. (14).

The degree of ellipticity is a second-order effect and cannot be readily approximated in the general case. For $\sigma_0/\omega\ll\epsilon'$, however, we can approximately set [see Eq. (10)]

$$\beta_+-\beta_-=\frac{1}{2}\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\left\{\frac{\sigma_+}{\sqrt{K_+'}}-\frac{\sigma_-}{\sqrt{K_-'}}\right\}. \quad 18)$$

Again letting $\sqrt{K_+'}=\sqrt{K_-'}=\sqrt{K'}$ in the denominator we have for the degree of ellipticity

$$\mathcal{E}=\frac{1}{4}\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\frac{1}{\sqrt{K'}}(\sigma_+-\sigma_-)l, \quad (19)$$

which with the aid of Eq. (5) reduces to (if $\omega\tau$ and $\mu B_z\ll1$)

$$\mathcal{E}=\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\frac{\sigma_0(\mu B_z)(\omega\tau)}{\sqrt{K'}}l. \quad (20)$$

Hence, for small losses the ellipticity depends linearly on the relaxation time.

If $\epsilon''>\epsilon'$ it is still possible to determine the relaxation time from a measurement of the ellipticity if the mobility has been previously determined from a measurement of the angle of rotation. It can also be shown that the angle of rotation reverses when the sign of the charge carriers or the direction of the magnetic field is reversed.

The whole derivation is based on the assumption that τ is energy independent. For more complicated cases where the relaxation time is a sensitive function of the velocity, the appropriate averages must be calculated in a manner similar to that proposed by Goldey and Brown[13] and Herring.[14]

Because of the small values of the quantities μB_z and $\omega\tau$ employed here, the cyclotron resonance condition $\omega\tau=\mu B_z$ cannot be expected to lead to any sharp maxima in the rotation vs B curve discussed by Suhl and Pearson.[5] The derivation also assumes the existence

[13] J. M. Goldey and S. C. Brown, Phys. Rev. **98**, 1761 (1955).
[14] C. Herring, Bell Syst. Tech. J. **34**, 237 (1955).

of only one effective carrier mass while in germanium the effective mass is anisotropic and more complex models of the energy surfaces apply in this case. For ellipsoidal energy surfaces, the Faraday rotation was discussed by Suhl and Pearson[7] and the equations developed by Lax and Roth[15] can also be applied to this problem.

In this case, the conductivity tensor in the coordinate system of each of the ellipsoids (in n-type germanium, 8 ellipsoids in the [111] direction) must be transformed to a coordinate system defined by the applied magnetic field and the transformed tensors for each of the ellipsoids added and divided by the number of ellipsoids, to obtain the conductivity tensor relevant to this problem. From this tensor a complex effective conductivity $[(\sigma_{eff})_\pm]$ can be calculated, which determines the propagation constants for right-handed and left-handed circularly polarized waves by the equation

$$k_\pm^2=\mu_0\omega^2\epsilon_{sl}'-i\mu_0\omega(\sigma_{eff})_\pm.$$

The real part of $[(\sigma_{eff})_\pm]$ then corresponds to the conductivities $\sigma_\pm$ and the imaginary part to the products $\omega(\Delta\epsilon)_\pm$ which were derived for the case of spherical energy surfaces in Eq. (5). It is, therefore, simple to determine the angle of rotation of the plane of polarization and ellipticity of the transmitted radiation if the appropriate values of $(\sigma_{eff})_\pm$ have been calculated for the case under consideration.

In general, the equations for the Faraday rotation and the ellipticity for ellipsoidal energy surfaces are much more complicated than for spherical energy surfaces and are dependent on the direction of the applied magnetic field. It can be shown however, that in n-type germanium for small magnetic fields applied in the [100] or [110] directions, and small values of $\omega\tau$,[16] the relation for the Faraday rotation derived for spherical energy surfaces [see Eq. (14)] applies to a first order of approximation also to ellipsoidal energy surfaces, if the mobility μ is interpreted as the Hall mobility (assuming energy independent relaxation times τ).

In n-type germanium, where there are 8 ellipsoidal energy surfaces in the [111] directions, the effective complex conductivity for a magnetic field applied in the [100] direction parallel to the direction of propagation of the electromagnetic wave is given by the relation[15]

$$(\sigma_{eff})_\pm=\sigma_\pm+i\omega(\Delta\epsilon)_\pm$$
$$=\sigma_0\left[1+i\left(\frac{m_1+2m_2}{2m_1+m_2}\right)\frac{qB}{m_2[(1/\tau)+i\omega]}\right]\Big/$$
$$1+\left(\frac{m_1+2m_2}{3m_1}\right)\frac{q^2B^2}{m_2^2[(1/\tau)+i\omega]^2}, \quad (21)$$

where m_1 and m_2 are the longitudinal and transverse components of the mass tensor. It is easy to show that this equation leads to the correct values for $\sigma_\pm$ and $(\Delta\epsilon)_\pm$ given by Eq. (5), if $m_1=m_2$.

For small values of the magnetic field, so that the second term in the denominator of Eq. (21) may be neglected, and small values of $\omega\tau$

$$(\Delta\epsilon)_+-(\Delta\epsilon)_-=2\sigma_0\left[\left(\frac{m_1+2m_2}{2m_1+m_2}\right)\frac{q\tau}{m_2}\right]B \quad (22)$$

leading to an expression for the angle of rotation equivalent to that given by Eq. (14), if

$$\mu=\left(\frac{m_1+2m_2}{2m_1+m_2}\right)\frac{q\tau}{m_2}. \quad (23)$$

But, for energy independent relaxation times, the ratio of the Hall mobility μ_H to the drift mobility μ is given by the relation[14]

$$\frac{\mu_H}{\mu}=\frac{3m_1(m_1+2m_2)}{(2m_1+m_2)^2}.$$

Hence,

$$\mu_H=\frac{3m_1(m_1+2m_2)}{(2m_1+m_2)^2}\cdot\frac{q\tau(2m_1+m_2)}{3m_1m_2}=\frac{m_1+2m_2}{2m_1+m_2}\frac{q\tau}{m_2}.$$

As far as the anisotropy of the effective masses is concerned, the electron mobility in germanium obtained from measurements of the Faraday rotation with small magnetic fields in the [100] direction and low frequencies or relaxation times (small values of $\omega\tau$) is, therefore, the Hall mobility.

In the case of n-type germanium, the complex effective conductivity $(\sigma_{eff})_\pm$ was also calculated for magnetic fields applied in the [110] direction (the direction in which the field was applied in our experiments) and, to the order of approximation discussed above, Eq. (14) was found to be valid in this case also.

For the case of holes in germanium the energy surfaces were, as a first approximation, assumed to be spherical, although they should be more accurately represented by warped spheres.[9]

III. CONNECTION WITH THE HALL EFFECT AND MAXWELL'S EQUATIONS

Making the assumption that the definition of the Hall field for the dc case

$$\mathbf{E}_H=R\mathbf{B}\times\mathbf{J}=\mu\mathbf{B}\times\mathbf{E},$$

where R is the Hall constant and $\mu=R\sigma_0$ (μ being the Hall mobility) is also valid for sinusoidally varying currents at low frequencies, the results of the Drude-Zener analysis, described in the previous section can, for the case of low frequencies, also be obtained by adding Hall-type fields to Maxwell's equations as

[15] Benjamin Lax and Laura M. Roth, Phys. Rev. **98**, 5, 549 (1955).

[16] The above approximations are valid in our experiments, where $\mu B\cong 0.06$ and $\omega\tau\cong 0.07$ (see Sec. V).

follows:

$$\nabla\times(\mathbf{E}+\mathbf{E}_H)=-\mu_0\frac{\partial\mathbf{H}}{\partial t},$$

$$\nabla\times\mathbf{H}=\sigma_0\mathbf{E}+\epsilon'\left[\frac{\partial\mathbf{E}}{\partial t}+\frac{\partial\mathbf{E}_H}{\partial t}\right].$$

For plane wave solutions of angular frequency ω, with the static magnetic field applied along the direction of propagation, the above equations lead to the following propagation constants for right- and left-handed circularly polarized light:

$$k_\pm^2=\mu_0\left\{\frac{\omega^2\epsilon'[1+(\mu B)^2]\mp\omega\sigma_0\mu B-i\omega\sigma_0}{1+(\mu B)^2}\right\}, \quad (24)$$

which are identical with the results derived from the Drude-Zener model at low frequencies ($\omega\tau\ll\mu B$).

The fact that the Hall field must not be added to the first term on the right-hand side of the second equation is because, in analogy with the dc case, the total current contains a displacement current, but no conduction current contribution from the Hall field.

By adding Hall-effect terms to Maxwell's equations Wicher[4] has predicted a Faraday rotation for artificial dielectrics. It can, however, be shown that if the dielectric constant is independent of $\mathbf{B}$ any current entirely in phase with $\mathbf{E}$ produces no rotation. If

$$\mathbf{J}=\alpha\frac{\partial\mathbf{E}}{\partial t}(\alpha \text{ real}),$$

and

$$\mathbf{E}_H=R\mathbf{B}\times\mathbf{J},$$

then

$$\nabla\times(\mathbf{E}+\mathbf{E}_H)=-\mu_0\frac{\partial\mathbf{H}}{\partial t}$$

$$\nabla\times\mathbf{H}=(\alpha+\epsilon_0)\left(\frac{\partial\mathbf{E}}{\partial t}+\frac{\partial\mathbf{E}_H}{\partial t}\right).$$

Assuming solutions identical to those giving Eq. (24)

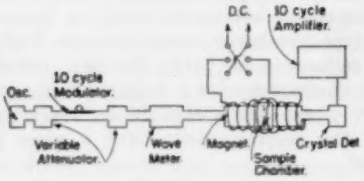
$$k_+^2=k_-^2=\omega^2\mu_0\epsilon'$$

and no Faraday rotation can be expected. The rotation predicted by Wicher[4] for artificial dielectrics would not occur if a Hall field term had been added to the left-hand side of the first of Maxwell's equations.

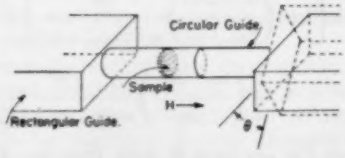
IV. EQUIPMENT AND MEASURING TECHNIQUES

A block diagram of the equipment appears in Fig. 1. The signal generator was a Klystron 723 *AB* tube and it was isolated from the rest of the circuit by two variable attenuators and one rotating attenuator, the latter modulating the power at 10 cycles per second so that a high-impedance 10-cycle amplifier (Electro-Mechanical Research Company) could be used in conjunction with the crystal detector. The wave guide was of the standard rectangular *X* band variety and transmitted. the standard TE_{10} mode except for the section containing the sample. This consisted of two lengths of cylindrical hollow brass pipe which fitted together inside a brass sleeve thus permitting the rotation of one of the circular guides relative to the other. With no sample in the guide the power in the rotating guide varied as $\cos^2\theta$, where θ is the angle between the principal planes of the two guides. With the sample inserted and no magnetic field applied this same $\cos^2\theta$ variation was observed except that there was a slightly larger minimum. In all cases the power ratio of maximum to minimum was 10^4 or more. The reflections from the junctions of the rectangular and circular guides should have no effect on the measured angle of rotation if they are isotropic. The $\cos^2\theta$ variation mentioned above indicates this to be the case. A solenoid slipped over the circular guide supplied the longitudinal magnetic field. The angle of rotation of the second guide relative to the first was measured by reflecting light from an attached mirror.

The sample was inserted in the nonrotating guide. It was found that the result did not depend on whether the sample made electrical contact (as determined by an ohm-meter) with the guide or not, although the spacing was not made more than about 0.5 mm. This is probably explained by the fact that the size of spacing and thickness of the sample (about 4.5 mm) makes propagation around the center impossible. The result was found to be unvaried when the sample was rotated about an axis along the direction of propagation illustrating the cubic nature of the single crystals used. An arbitrary zero ellipticity and rotation both independent of the magnetic field were introduced by inserting into the rotating guide a rectangular post of lucite inclined at an arbitrary angle relative to the original direction of polarization. This changed the original angle of the minimum by about 2.0° but the



(a)—Experimental Set-up.



(b)—Detail Drawing of Sample Chamber.

FIG. 1. Measuring equipment.

measured rotation caused by the field about this new minimum remained the same. These angles are of the order of 3.5°. When the attenuators isolating the oscillator were varied it was found that the rotation caused by the magnetic field was independent of the power delivered by the oscillator over a range of 10^2 in power. The angle was determined by taking about 6 points around the minimum and drawing a symmetric curve. The angle was found to reverse when the magnetic field was reversed. The measured angle was reproducible to within about 4 percent. The germanium utilized had a skin depth of about 3 mm so that with samples about 4.5 mm thick all problems of multiple transmission due to internal reflections and standing waves should be eliminated.

The analysis above has been made for a plane wave while the waves used in the experiment were propagating in the TE_{11} mode in the cylindrical guide. This problem is covered by Suhl and Walker,[17] but to apply their solution to the present case of lossy materials would involve complex trial and error solutions of Bessel functions of complex arguments. We here present a semiquantitative analysis. While the guide mode differs from a plane wave by having (a) a longitudinal component of the magnetic field, (b) H not constant in a transverse plane, (c) E not constant in a transverse plane, it was felt that only the absence of a uniform electric field in the transverse plane would cause any important difference between the two solutions.

A diagram of the electric field is given in Fig. 2.[18] From this figure, it appears that the actual field configuration in the guide differs from that of a plane polarized wave by the decrease of the field vector in the plane of polarization. The former does not lead to a change of the angle of rotation as the wave passes through the sample, but only to a change in the amplitude or intensity of the field component parallel to the major axis of the ellipse. Although a field component at right angles to the plane of polarization will lead to a change of the angle of rotation compared with the plane polarized incident wave, it is easily seen from symmetry considerations that the contributions of this field component are equal and opposite at equal and opposite distances from the center of the guide and will, therefore, on the average, cancel out.

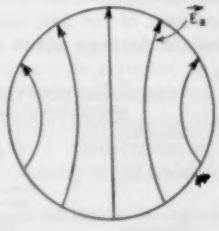


FIG. 2. Field configuration in circular guide (TE_{11} mode).

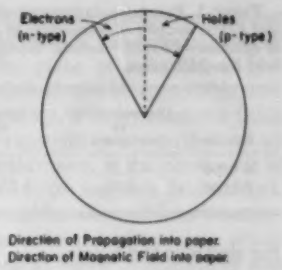


FIG. 3. Rotation of the plane of polarization for n and p type semiconductors.

We may, therefore, conclude that the plane wave solutions which we have used in our theoretical considerations are good approximations in determining the angle of rotation of the plane of polarization. The degree of ellipticity of the transmitted radiation, however, will be influenced by deviations from the assumption of strictly plane polarized incident waves.

V. RESULTS

Measurements were made on two single crystals of germanium, which were made available to us through the courtesy of the Philco Radio Corporation. The crystals were large enough to fill the entire cross section of the guide. One sample was n-type with a resistivity of 16 ohm-cm and the other p-type with a resistivity of 19.8 ohm-cm. The magnetic field was applied in the [110] direction. The observed sense of rotation was in agreement with the theoretical arguments given above. (Fig. 3.)

The experimental results for these two samples at room temperature are shown in Table I.

The effective Hall mobility was calculated from Eq. (14) using a value of 16 for the static dielectric constant K_{sl}' and a value of 1.3×10^{-12} sec for the relaxation time τ. An approximate value for the effective Hall mobility can be obtained from Eq. (15) which does not contain the relaxation time. From Eq. (15) we obtain values of 3720 cm²/volt-sec and 3080 cm²/volt-sec for the effective Hall mobilities of the n- and p-type samples respectively. These values differ by only *ca* 2 percent from the more accurate values obtained from Eq. (14), showing that the dependence of the Faraday rotation on the relaxation time is not very large.

From the values for the effective Hall mobility of these two samples, it is easy to derive the Hall mobility of electrons and holes in germanium at room temperature.

An elementary analysis gives the following relations between the effective Hall mobilities $(\mu_{eH})_n$ and $(\mu_{eH})_p$ of the n- and p-type samples, respectively, obtained from our experiments, and the Hall mobilities μ_e and μ_h of

[17] H. Suhl and L. R. Walker, Phys. Rev. 86, 122 (1952).
[18] N. Marcuvitz, *Wave Guide Handbook* (McGraw-Hill Book Company, Inc., 1951), Vol. 10, Radiation Laboratory Series, p. 71. This reference also contains an analytic expression for the field (p. 69).

TABLE I. Faraday rotation and effective Hall mobilities of n- and p-type samples of germanium at room temperature. Thickness of samples $=4.6\times10^{-4}$ meters. Wavelength $=3.434$ cm. Magnetic field $B_z=1430$ gauss.

Sample	Dc conductivity σ_0 (ohm-meter)$^{-1}$	Measured angle θ	Effective Hall mobility μ_{eff} cm^2/volt-sec
n-type	6.25	$-3°28'\pm4'$	3640 ± 70
p-type	5.05	$+2°22'\pm4'$	3040 ± 85

the electrons and holes:

$$(\mu_{eff})_n=\frac{\mu_e}{b}\left\{\frac{bB(n_i/n_h)^2-1}{B(n_i/n_h)^2+1}\right\},$$

$$(\mu_{eff})_p=\mu_h\left\{\frac{1-(n_i/n_h)^2Bb}{1+(n_i/n_h)^2B}\right\}.$$

Here n_h is the hole concentration in the sample, b and B are the Hall and drift mobility ratios respectively, while n_i is defined by the relation $n_en_h=n_i^2$. (n_e being the electron concentration.)

The ratio n_i/n_h can be obtained from the appropriate roots of the equation

$$\frac{\rho_{n\text{ or }p}}{\rho_i}=\frac{n_i/n_h(B+1)}{1+(n_i/n_h)^2B},$$

where $\rho_{n\text{ or }p}$ are the dc resistivities of our samples and ρ_i is the intrinsic resistivity of germanium, which was taken to be 47 ohm-cm.[19]

For the drift mobility ratio a value of 2.1 was chosen[19] while the Hall mobility ratio was taken to be 1.18 in agreement with the measurements of Morin[20] as well as our own results.

The Hall mobility values of the electrons and holes in germanium at room temperature thus obtained are given in Table II and are compared with those of Morin.[20]

The errors in the mobility values as stated in Table II apply only to the precision with which the angles of rotation could be determined. The relatively good agreement between our values and those of Morin is surprising considering the rather rough nature of our experiments and the many approximation inherent in our theoretical considerations and must be considered accidental.

Measurements of the ellipticity δ of the transmitted radiation in these samples gave values of the correct order of magnitude. The latter were, however, too small to permit an accurate determination of the relaxation time, since the small elliptical component always present in the incident radiation in the guide was of the same order of magnitude as that produced by the transmission through the sample. This background ellipticity can be caused by imperfections in the circular guide or nonuniformity of the sample. Furthermore, the difference in the power received by the detector between radiation along the major and minor axes of the ellipse was too large to be measured accurately with our equipment. To obtain a value of the relaxation time, the tests will have to be carried out at lower temperatures or with somewhat larger magnetic fields. In any case, the anisotropy of the effective masses of the charge carriers will have to be taken into account even for weak magnetic fields, since the ellipticity is a second order effect (see Sec. II). In addition, the fact that the incident radiation in the guide is not strictly plane parallel will have a far larger influence on the ellipticity than on the angle of rotation and the applicability of the equations developed in Sec. II to this problem requires careful consideration (see Sec. IV).

VI. APPLICATION TO POWDERS

The measurements described in the previous sections can also be applied to powders. This method has the attractive feature that no electrodes need to be applied to the specimen.

In order to apply the technique, it is necessary to find the field existing within the powder particle in terms of the average field in the wave guide. A procedure which can be applied to specimens in which the displacement current is large in comparison with the conduction current has been worked out by Smith.[21] He gives a formula for the average field within randomly oriented particles of different shapes in terms of the microscopic field which is averaged over both particles and the space between them.

In the case of spheres, his formulas are the same as those obtained by the Clausius-Mosotti procedure but differ markedly when particles of other shapes such as prolate spheroids are considered.

Some preliminary measurements have been made on a germanium powder obtained by crushing in a mortar and pestle some single crystal germanium of high quality. The rotation observed was much less than that obtained from the same amount of crystalline material. The reason for this is that the maximum density obtainable is only about 50 percent of the single crystal value. Germanium has a high dielectric constant and hence most of the field appears in the interstices between the particles rather than in the particles them-

TABLE II. Hall mobilities of electrons and holes in germanium at room temperature.

	Hall mobility (cm^2/volt-sec)	
	Faraday rotation measurements	Hall effect measurements (Morin)
Electrons	3780 ± 70	3900
Holes	3330 ± 85	3300

[19] L. P. Hunter, Phys. Rev. **91**, 579 (1953).
[20] F. Morin, Phys. Rev. **93**, 62 (1954).
[21] R. S. Smith, Thesis, University of Pennsylvania (1955). See also Tech. Rept. 12, Contract N6-onr-24914, January 1, 1955.

selves. The effect did appear to be of the right order of magnitude, but no quantitative measurements were attempted because of the unknown particle shape distribution of our crushed specimen.

Further work on powders is planned for the immediate future.

VII. DISCUSSION

A measurement of the rotation of the plane of polarization when a plane polarized electromagnetic wave passes through a nonferromagnetic semiconductor and a small magnetic field is applied along the direction of propagation yields to a first order of approximation values of the product $\sigma_0\mu/\sqrt{K_{ii}'}$ where σ_0 is the dc conductivity, μ the Hall mobility of the free charge carriers and K_{ii}' the static dielectric constant of the material [see Eq. (15)]. Hence, if the dc conductivity and the static dielectric constant are known, the mobility can be determined by this technique. Conversely, K_{ii}' can be obtained if the dc conductivity and the Hall mobility of the sample are known. The method has the advantage over dielectric measurements in that measurements of the angle of rotation of the plane of polarization are considerably easier than measurements of standing wave ratios. Furthermore, the method should be applicable to measurements on powdered samples without electrodes. Since the angle of rotation depends only in the second order of approximation on the relaxation time through the factor $\sigma\tau/\epsilon_0$ and since this is small for almost all semiconductors, even at microwave frequencies,[22] an accurate knowledge of the relaxation time of the charge carriers is not essential in order to obtain accurate values of the Hall mobility for most semiconductors.

For weak magnetic fields (at least when these are applied in the [100] or [110] directions) and small values of $\omega\tau$, the equations governing the Faraday rotation are, to a first approximation, the same for ellipsoidal or spherical energy surfaces, and, by the use of Eq. (14), measurements of the above quantity will lead to a determination of the Hall mobility of the charge carriers. In general, the angle of rotation may depend on the direction of the applied magnetic field because of the nonspherical nature of the energy surfaces.

Although a measurement of the angle of rotation of the plane of polarization cannot be expected to give accurate information on the relaxation time, the relaxation time and also the effective masses of the charge carriers can be obtained from a measurement of the phase shift or the degree of ellipticity of the transmitted radiation, if the mobility of the charge carriers has already been determined from a measurement of the angle of rotation of the plane of polarization [see Eq. (20)].[23] The phase shift should be appreciable, and could be measured accurately, if larger magnetic fields are employed or if the experiments are carried out at lower temperatures. Furthermore, if the phase shift is investigated as a function of the magnetic field strength B, the relaxation time could be obtained from measurements of small differences in the phase shift which should be detectable even at higher temperatures and low frequencies (small values of $\omega\tau$). Since the phase shift is a second order effect the anisotropy of the effective masses must be taken into account and the simple Eq. (20) is no longer valid (see Sec. II). Hence, the ellipticity is a function of the direction of the applied magnetic field. In principle, then, the effective masses of the charge carriers may be determined from measurements of the Faraday rotation and the phase shift as a function of the orientation of the applied magnetic field, as was attempted in the experiments of Suhl and Pearson.[5] While this method would be more inaccurate than the cyclotron resonance technique, it may be applicable to measurements at somewhat higher temperatures than the latter and to substances in which the cyclotron resonance method cannot be used.

For general applications, the theory has to be worked out in detail in cases where the relaxation time is a function of the electron energy and direction of motion, but this difficulty would also be experienced in the interpretation of dielectric measurements although the mathematical formalism may be simpler in the latter case.[24]

ACKNOWLEDGMENTS

The authors are greatly indebted to Professor Park H. Miller, Jr. for suggesting the problem and for helpful advice and encouragement throughout the course of this investigation. Our thanks are due to H. Schultz for computational assistance and to the Philco Radio Corporation for supplying single crystals of germanium. We are also greatly indebted to Dr. G. L. Pearson of the Bell Telephone Laboratories and to Dr. B. Lax of the Lincoln Laboratories, Massachusetts Institute of Technology for submitting to us unpublished manuscripts and manuscripts prior to publication.

[22] R. Rau, Tech. Rept. 2, Contract AF 33(616)-78, January 15, 1955.

[23] For materials of small conductivity and spherical energy surfaces the phase shift depends linearly on the relaxation time.

[24] The mobility obtained from the Faraday rotation can probably not be interpreted simply as the effective Hall mobility of the charge carriers, even for weak magnetic fields, if the energy dependence of the relaxation time is taken into account, although such a interpretation seems correct as far as the anisotropy of the effective masses is concerned. (See reference 14.)

Schottky Effect for SrO Films on Molybdenum*

G. A. Haas† and E. A. Coomes
University of Notre Dame, Notre Dame, Indiana

(Received June 3, 1955; revised manuscript received July 29, 1955)

The Schottky effect and the periodic deviations about the Schottky line have been simultaneously investigated for thin films of SrO on molybdenum. The experimentally determined Schottky slope is in good agreement with the theoretical value of e^3/kT for deposits less than a monolayer; but becomes larger than theory for more than a monolayer. The periodic deviations obtained for a clean molybdenum surface wash out when as little as 1/60 monolayer of SrO is deposited on the surface. The implication is that the electron reflection coefficient near a Mo-SrO surface cannot be associated with the simple μ reflection found for clean metals.

THE Schottky effect and the periodic deviations from the Schottky line have been suggested[1] as a tool for exploring the nature of the surface potential barrier. The thermionic emission in the Schottky region for thin films of SrO on molybdenum has been investigated experimentally, using a cylindrical anode-guard ring structure coaxially mounted with a 0.003-inch polished molybdenum wire. Strontium oxide was deposited by evaporation from four coated platinum source filaments[2] using the deposition techniques of Moore and Allison[3] and the emission measurement methods developed for periodic deviation studies.[4] Measurements were carried out up to fields of 10^5 volt cm^{-1}. Table I gives the slope of the experimental Schottky line as compared with the theoretical slope e^3/kT, for various deposits of SrO. Calibrations of the source filaments have shown that it required approximately 8 hours of evaporation to deposit the equivalent of one uniform monomolecular layer on the receiver filament. Within the accuracy of temperature determinations, the experimental and theoretical Schottky slopes are in agreement for less than a monolayer. For as little as 1/60 monolayer (7 minutes deposit by evaporation) the thermionic emission increased by a factor of 10^3 over that for clean molybdenum, indicating that the electron current observed came from regions covered by SrO. That the Schottky law was obeyed gives indication of a mirror image potential in the λ region[5] of the barrier, for SrO deposits of less than a monolayer.

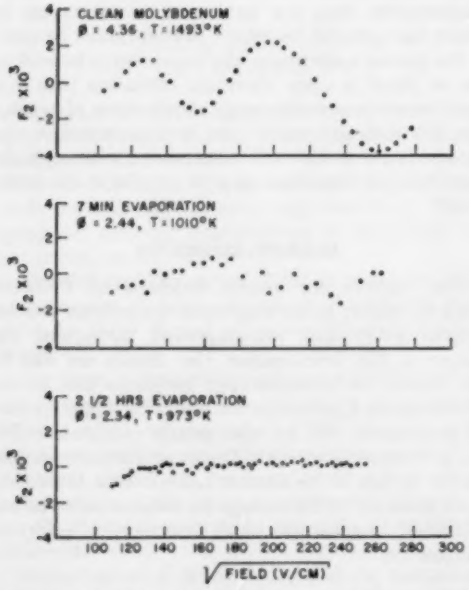

CLEAN MOLYBDENUM
ϕ = 4.36, T = 1493°K
7 MIN EVAPORATION
ϕ = 2.44, T = 1010°K
2 1/2 HRS EVAPORATION
ϕ = 2.34, T = 973°K
F2 X10³
100 120 140 160 180 200 220 240 260 280 300
√FIELD (V/CM)

FIG. 1. Effect on the Schottky deviations as a thin film of SrO is deposited on a clean molybdenum filament. F_2 represents the periodic deviation from the Schottky line. See reference 3.

For coverages greater than a monolayer, the experimental slope becomes larger than the theoretical slope; it is nearly twice as large at approximately 16 layers. This effect is common in the study of oxide cathodes where there is also more than a monolayer of alkaline earth oxide present. An explanation has been suggested

TABLE I. The variation in Schottky slope for a molybdenum filament on which SrO has been deposited as a thin film. The equivalent of one monomolecular layer of SrO is deposited during 8 hours of evaporation. The theoretical slope is e^3/kT.

Evaporation time	Work function	Experimental slope / Theoretical slope
0 (clean Mo)	4.36 ev	1.03±0.03[a]
7 minutes	2.44±0.04	0.99
15 minutes	2.34±0.04	1.03±0.03[b]
2.5 hours	2.34±0.04	1.00
3 hours	2.34±0.04	0.99±0.01[c]
8 hours	2.6 ±0.1	1.21−1.45
130 hours	2.7 ±0.1	1.61−1.92

[a] Five runs.
[b] Three evaporations.
[c] Two evaporations.

* Supported in part by the Office of Naval Research.
† Now at the U. S. Naval Research Laboratory, Washington, D. C.
[1] C. Herring and M. H. Nichols, Revs. Modern Phys. **21**, 185 (1949).
[2] The source filaments, prepared by spraying a coating of SrO in a nitrocellulose binder on highly refined platinum, were furnished by G. E. Moore, Bell Telephone Laboratories, Murray Hill, New Jersey.
[3] G. E. Moore and H. W. Allison, Phys. Rev. **77**, 246 (1950).
[4] Munick, LaBerge, and Coomes, Phys. Rev. **80**, 887 (1950).
[5] Juenker, Colladay, and Coomes, Phys. Rev. **90**, 772 (1953).

by Nergaard[6] for the case of films thick enough to exhibit semiconducting properties.

The effect on the periodic deviations[5] from the Schottky line resulting from a deposition of SrO on molybdenum is illustrated in Fig. 1. There is a washing out of the deviations for deposits as small as 1/60 monolayer, and at approximately ½ monolayer there are no deviations from the Schottky line greater than 1/100%. Since the surface barrier is mirror image in the λ region, as suggested above, these data would indicate that with the SrO film, electron reflections near the emitter surface cannot be associated with the simple μ coefficient[6] found for clean metals.

[6] L. S. Nergaard, RCA Rev. 18, 464 (1952).

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Search for the Hall Effect in a Superconductor. II. Theory

H. W. Lewis
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received July 15, 1955)

Various theoretical explanations for the absence of a Hall effect in a superconductor are considered, and none are found convincing. It is concluded that this property lies outside the existing body of theory, and the general requirements it imposes on a future theory are adduced.

I. ORIENTATION

IN a previous work[1] the early inconclusive result of Kamerlingh Onnes, that the Hall coefficient of a metal vanishes when it goes into the superconducting state, was confirmed for a vanadium sample. (In fact, it was shown that the Hall coefficient was reduced to less than about 20% of its value in the normal state. It would probably be more difficult to explain its reduction to a finite smaller value than to explain its vanishing completely, so that, for the purposes of the following, we will suppose that these experiments demonstrate the disappearance of the Hall effect in the superconducting state.) It is the purpose of this note to discuss the theoretical implications, if any, of this fact.

To begin with, it is probably well to dispose of some previous arguments that have been advanced, to the effect that it is perfectly obvious that there ought to be no Hall effect in a superconductor. These arguments take several forms, and we discuss them in order.

First, it has been argued that, since a magnetic field can hardly penetrate a superconductor, and, since a magnetic field is essential to a Hall effect, such an effect is precluded. In fact, a magnetic field penetrates a superconductor to a known depth of about 10^{-5} cm, and within this penetration depth the current densities can be quite large. This property of a superconductor is fully taken into account in the earlier paper, in evaluating the magnitude of effect to be expected.

The second argument is due to Pippard[2] and is as follows: if there were a Hall effect in a superconductor it would be possible, in a static magnetic field, to impose a resistive load on the Hall voltage, and so to draw energy from the field. This would, however, cause the collapse of the magnetic field, which is in contradiction to the observed stability of the current-carrying state in a superconductor. Consequently, there must be a contact potential at the surface, which is just sufficient to cancel the Hall voltage, and to render it unobservable.

The answer to this comes in two parts. (We must grant, to begin with, that the gedanken-experiment would go as described, though it has not been tried. This is very probable.) In the first place, one must consider the magnitudes involved in this gedanken-experiment. A superconducting sphere of radius a, in an applied magnetic field H, has an energy excess over a nonsuperconducting sphere equal to $3H^2a^3/8$. The Hall voltage[1] between the pole and equator would be $9RH^2/32\pi$, where R is the Hall coefficient, and, as before, H is the applied field at infinity. If a resistive load Z is now applied, the decay time T will be given by $T=128\pi^2Za^3/27R^2H^2$. Suppose $a\approx1$ cm, $Z\approx1$ ohm $=10^9$ emu, $H=100$ gauss, $R=100\times10^{-6}$ emu (tin or lead); then $T\approx5\times10^{14}$ sec. Thus, the field collapse of which we have spoken above would take something more than fifteen million years. The stability of the super currents can hardly be considered to be established over such periods of time. This time can, of course, be somewhat reduced by a suitable choice of different numbers from those used above, but cannot be brought down to the experimental time, which is of the order of weeks.

Thus, one is finally forced to make this argument a matter of principle, and here the question is less clear. If one accepts the idea that the field collapse will indeed occur over such a long period of time, one accepts with it the concept of a lower energy state of a supercon-

[1] H. W. Lewis, Phys. Rev. 92, 1149 (1953).
[2] Unpublished, but quoted in D. Schoenberg, *Superconductivity* (Cambridge University Press, London, 1952), p. 49. I have had the pleasure of an interesting conversation with Professor Pippard on this question.

ductor, in which the field is no longer excluded. Thus the normal superconducting state, exhibiting the Meissner effect, must be regarded as metastable, rather than absolutely stable. This concept is distasteful to some physicists, but cannot be regarded as in any sense excluded by the information at hand. The answer to such a question must probably ultimately come from a satisfactory theory of the superconducting state.

The third argument we will discuss is given by London[3] and is substantially as follows, if we confine ourselves, for the moment, to the discussion of the forces acting on the superelectrons, and ignore the normal electrons.

Consider the Lorentz expression for the force per unit volume acting on the former

$$\mathbf{F}=\rho_s[\mathbf{E}+(1/c)\mathbf{v}_s\times\mathbf{H}], \tag{1}$$

where ρ_s is the charge density of the superelectrons, $\mathbf{v}_s$ their velocity, c the speed of light, and $\mathbf{E}$ and $\mathbf{H}$ the electric and magnetic fields. This can be transformed, by using the London equations

$$\operatorname{curl}(\Lambda\mathbf{j}_s)=\operatorname{curl}(\Lambda\rho_s\mathbf{v}_s)=-(1/c)\mathbf{H}, \tag{2}$$

$$\Lambda(\partial/\partial t)(\mathbf{j}_s)=\mathbf{E} \tag{3}$$

to eliminate the field strengths. Λ is a constant. Then (1) can be written after suitable transformation

$$\mathbf{F}=\frac{\partial}{\partial t}(\Lambda\rho_s\mathbf{j}_s)+\operatorname{div}(\Lambda\mathbf{S}), \tag{4}$$

where $\mathbf{S}$ is a tensor defined by

$$S_{\mu\nu}=j_\mu j_\nu-\tfrac{1}{2}j^2\delta_{\mu\nu}, \tag{5}$$

and which represents the kinetic stress tensor of the supercurrents. Thus (4) tells us that the momentum delivered to the supercurrents, per unit volume, and per unit time, is divided between an increase of the momentum density in the volume, and a transport of stress across the boundary of the region considered, $\Lambda S_{\mu\nu}$ representing the flow in the ν direction of the μ component of momentum density. If we were to add the stress tensor of the electromagnetic field we would find an over-all conservation law of the form

$$\partial\mathbf{P}/\partial t+\operatorname{div}(\mathbf{T})=0, \tag{6}$$

where we have derived above the expressions for the contributions of the superelectrons to $\mathbf{P}$ and $\mathbf{T}$. The contributions due to the field have the standard form.

It is now argued that, in the stationary case, the first term on the right side of (4) vanishes, as does $\mathbf{E}$, according to (3). Thus, the second term on the right side of (1), representing the magnetic force on the superelectrons, is converted completely into the kinetic stress of the supercurrents, represented by $\mathbf{S}$ in (4). We have thus an equilibrium situation, in which the magnetic forces are balanced by the inertial forces, without an electric field, hence without a Hall effect. At the surface of the material, where $\mathbf{S}$ changes discontinuously, becoming zero outside the metal, there must be surface forces which finally take up the transferred momentum, and which are responsible for the fact that even a superconducting wire can be used in an electric motor. This, then, is a coherent picture, in which the Hall effect has no place.

To study it more carefully, we must go back to the London equations (2) and (3), and to their basis. Here one can adopt one of two points of view; one can postulate the equations in the form given, or one can try to justify them, or at least make them plausible, by studying the motion of a frictionless electronic fluid in an electromagnetic field. London has done this in the book previously referred to.[3] The motion of each electron, under the force represented by (1), will be determined by

$$m(d\mathbf{v}_s/dt)=e(\mathbf{E}+(1/c)\mathbf{v}_s\times\mathbf{H}), \tag{7}$$

where m and e are the mass and charge of the electron. Further, the acceleration can be transformed into

$$\frac{d\mathbf{v}_s}{dt}=\frac{\partial\mathbf{v}_s}{\partial t}+(\mathbf{v}_s\cdot\boldsymbol{\nabla})\mathbf{v}_s=\frac{\partial\mathbf{v}_s}{\partial t}+\tfrac{1}{2}\boldsymbol{\nabla}(\mathbf{v}_s^2)-\mathbf{v}_s\times\operatorname{curl}\mathbf{v}_s, \tag{8}$$

so that (7) can be written

$$\frac{\partial\mathbf{v}_s}{\partial t}+\tfrac{1}{2}\boldsymbol{\nabla}(\mathbf{v}_s^2)-\frac{e}{m}\mathbf{E}=\mathbf{v}_s\times\left[\operatorname{curl}\mathbf{v}_s+\frac{e}{mc}\mathbf{H}\right]. \tag{9}$$

If we now take the curl of both sides, call the expression in the square bracket $\boldsymbol{\Omega}$, and use one of Maxwell's equations, it is easy to see that we have

$$\frac{\partial\boldsymbol{\Omega}}{\partial t}=\operatorname{curl}(\mathbf{v}_s\times\boldsymbol{\Omega}). \tag{10}$$

Consequently, if $\boldsymbol{\Omega}$ is initially zero everywhere, it will remain zero at all times. It is thus consistent to postulate, in addition to the equations of motion, the equation

$$\boldsymbol{\Omega}=0 \tag{11}$$

It is obvious that this is identical with London's first equation (2), if we identify Λ with $m/e\rho_s$. This is in fact the right order of magnitude for Λ, if we take for ρ_s about one electron per atom.

Thus, one is led to London's equation (2), with a meaning for the constant Λ. Now combine this with (9), to obtain

$$\partial\mathbf{v}_s/\partial t=(e/m)\mathbf{E}-\tfrac{1}{2}\boldsymbol{\nabla}(\mathbf{v}_s^2), \tag{12}$$

which is, apart from the last term, identical with (3), and with the same value of Λ. London now argues that, although (12) is probably a more consistent expression

[3] F. London, *Superfluids* (John Wiley and Sons, Inc., New York, 1950), p. 70.

of the equations for a superconductor than is (3), the gradient term is always small, and (3) is simpler. We will argue that the extra term radically alters the conclusion drawn above about the absence of the Hall effect.

Indeed, if one goes through the argument contained between Eqs. (1) and (6), but using (12) instead of (3), one finds that it goes exactly as before, except that the tensor **S** is replaced by a new tensor **S′**, lacking the last term in (5), and therefore given by

$$S_{\mu\nu}' = j_\mu j_\nu. \tag{13}$$

Also, it is no longer the case that $E=0$ in a stationary state, but, from (12)

$$\mathbf{E} = \frac{m}{2e}\nabla(\mathbf{v}_s^2) = \frac{e}{2m}\nabla(\Lambda^2\mathbf{j}_s^2), \tag{14}$$

since $m\mathbf{v}_s = e\Lambda\mathbf{j}_s$, with the value adopted for Λ in this picture.[4]

There are therefore two points in which the conclusions differ from those drawn by London on the basis of (2) and (3). First, the expression for the kinetic stress tensor is changed from (5) to (13), thus eliminating the Bernouilli pressure term. Thus, since the normal components of j must be zero at the surface, the discontinuity in stress across the surface no longer exists, and the same is true of the surface forces which appear in the original picture. Secondly, the electric field is no longer zero, but is given by (14), and constitutes the Hall field. The usefulness of the superconducting wire in an electric motor is still guaranteed, but the forces communicating with the surface are electric forces operating on the metallic lattice. This is the more usual situation.

To spell out this point, consider the following particularly simple stationary situation. A free plane surface of a superconducting body has a magnetic field H_0 parallel to the body just outside it, and currents j inside. Call the interior normal the z direction, the magnetic field direction y, and the current direction x, supposing them to be mutually perpendicular. Then, according to Eqs. (2) or (11), combined with Ampere's law, for the stationary case,

$$\mathrm{curl}\mathbf{H} = (4\pi/c)\mathbf{j}, \tag{15}$$

we can find that the magnetic field as well as the current decrease exponentially going into the material as

$$\begin{aligned} \mathbf{j} &= \mathbf{j}_0 \exp[-z/\lambda], \\ \mathbf{H} &= \mathbf{H}_0 \exp[-z/\lambda], \end{aligned} \tag{16}$$

where the penetration depth λ is given by

$$\lambda^2 = \Lambda c^2/4\pi = mc^2/4\pi e\rho_s. \tag{17}$$

[4] Substantially this conclusion was reached by F. Bopp, Z. Physik **107**, 623 (1937), in a discussion of the derivation of London's equations from the acceleration theory.

With this distribution of field and current we note that $(\mathbf{v}_s\cdot\nabla)\mathbf{v}_s = 0$, so that there are no so-called "inertial forces." This is true in both representations of the London equations. Indeed, the "inertial forces" in the version given by Eqs. (2) through (5) can be regarded as arising in breaking up this relation into

$$(\mathbf{v}_s\cdot\nabla)\mathbf{v}_s = \tfrac{1}{2}\nabla(\mathbf{v}_s^2) - \mathbf{v}_s\times\mathrm{curl}\mathbf{v}_s, \tag{18}$$

and then ignoring the first term on the right-hand side, as negligible. In fact, in the case here presented, it is equal and opposite to the last term, which is usually kept, and the electric field given by (14) is easily seen to be equal to the usual Hall field. Since the net force on the electrons is now zero, it is clear that the magnetic force has been turned into electric forces that act on the metallic lattice.

In conclusion, then, we may say that there are available two forms of London's equations, (2) and (3), and (11) and (12), respectively. We may denote them by I and II. They differ by terms which are usually negligible, but which are of utmost interest here. I is the set originally proposed by London, and leads to the prediction that the Hall effect is absent. II is the set that appears naturally from a study of the dynamics of a frictionless electronic fluid, and leads to a normal Hall effect. II is probably, according to London, a somewhat more consistent expression of the electrodynamics of a superconductor, if we believe that frictionless motion of electrons is in any way involved. It is probably also worth noting that the equations II constitute, to second order, a relativistically covariant set, whereas the equations I do not. At the very least, it appears that from this type of argument one cannot predict whether or not a superconductor ought to exhibit a Hall effect.

Note added in proof.—Professor Bardeen has kindly called to my attention an article by J. Lindhard [Phil. Mag. **44**, 916 (1953)] in which it is argued that the London equations do not represent a limiting case of a frictionless electronic fluid. This argument depends upon the zero-point velocity of the electrons in a metal, and must be regarded as rendering still more mysterious the origin of the London equations.

In all the above we have ignored the role of the "normal" electrons in the superconductor. The next sections will be devoted to a study of their influence.

II. THE TWO-FLUID MODEL

We want here to set out briefly those characteristics of the "two-fluid" model[5] which will be essential to the subsequent discussion. The model has been developed in much greater specificity than we need for our purposes. Indeed, for these off-equilibrium situations the specific forms of the model that are currently in vogue cannot be considered well established.

[5] Gorterand Casimir, Physik. Z. **35**, 963 (1934); Z. tech. Phys. **15**, 539 (1934); see also reference 2, pp. 194 ff.

We consider the electrons to divide themselves into two groups, with the relative populations of these groups determined by a parameter f that can vary between zero and unity, and which we call the "fraction of normal electrons." These two groups are supposed to coexist in the same volume, and to be in thermal, but not mechanical interaction. Thus they have, in local equilibrium, the same temperature T and Gibbs free energy density μ (per unit mass) but not necessarily the same pressures.[6,7] It follows that in transport equilibrium we have

$$\begin{aligned}\nabla p_n &= \rho_n \nabla\mu + \sigma_n \nabla T,\\ \nabla p_s &= \rho_s \nabla\mu + \sigma_s \nabla T,\end{aligned} \tag{19}$$

where p, ρ, and σ are the pressure, density, and transport entropy per unit volume, and the subscripts refer to normal and superconducting components. These equations determine the thermomechanical forces on the two groups of electrons in terms of the gradients of the macroscopic thermodynamic quantities. We will suppose that the transport entropy of the superconducting electrons is zero, i.e., $\sigma_s=0$.

If Eqs. (19) were coupled with suppositions about the kinetics of entropy transport in the electron fluid, we would have a complete scheme for treating the dynamical properties of the two fluid model. In the case of liquid helium II, where analogous considerations apply, they lead to the phenomena of second sound, fountain effect, anomalous heat conductivity, etc. All these are associated with relative motions of the two components, which bring the thermomechanical forces into play.

In the case of superconductors, however, the friction involved in the mass motion of the normal electrons serves to destroy some of these effects. For example, one might expect, in a superconductor, an anomalous heat conductivity produced by the counterflow of the two components, as in helium. However,[6] the resistance to the motion of the normal electrons reduces this to an unobservable fraction of the normal heat conductivity, and it has never been observed. Similarly for the "second sound" phenomenon in a superconductor. Here the resistance can be expected to attenuate the wave in much less than one wavelength, and it too has never been observed. The same argument applies to the absence of thermoeffects in a superconductor. Clearly, if one is to observe the effects of (19) in a superconductor, one must do so without moving the normal electrons.

The Hall effect situation is one in which currents are carried by the superelectrons alone, so that only they experience the magnetic forces. A complete discussion of the behavior of the electrons in the skin must therefore take into account the details of the two-fluid model. This is the subject of the next section.

III. THE HALL EFFECT IN THE TWO-FLUID MODEL

In the presence of electric and magnetic fields the forces on both the normal and superconducting electrons are given by combining (19) with the Lorentz force. If we suppose that only the superelectrons are moving (with a mean velocity v_s), we have, for the forces on the electrons (per unit volume)

$$\begin{aligned}\mathbf{F}_n &= \rho_n \nabla\mu_e + \sigma_n \nabla T,\\ \mathbf{F}_s &= \rho_s \nabla\mu_e + (e/mc)\rho_s[\mathbf{v}_s \times \mathbf{H}],\end{aligned} \tag{20}$$

where we have introduced the electrochemical potential μ_e to include the electric forces. Equilibrium is determined by the absence of a net force on the superelectrons, so that

$$\nabla\mu_e + (e/mc)[\mathbf{v}_s \times \mathbf{H}] = 0. \tag{21}$$

We have now to consider the normal electrons. If there is a net force on them, it will result in an internal circulation of current, and therefore a dissipative effect.[8] Within this framework, the only way this can be avoided is for the thermoelectric force, represented by the term in the gradient of the temperature in (20) to cancel the electrochemical force—so that we would have a thermal gradient in the presence of a magnetic field. This would amount to substituting an Ettingshausen effect for the Hall effect, and is doubtless what would happen if the metal were a thermal insulator. It is not hard to see, however, that the actual thermal gradient will be severely reduced by heat transport in the metal, and, in fact, by a factor of approximately $(kT/E_F)^2$ which is of order 10^{-3}. This can be estimated by writing down the entropy flow balance in the steady-state condition. Thus, the metal can be considered to be isothermal, and a net force equal to $\rho_n \nabla\mu_e$ must be supposed to act on the normal electrons. Thus we have both a dissipative normal current, and a Hall effect.

Deferring for a moment discussion of the fact that the two-fluid model fails to account for the absence of the Hall effect, let us consider first the magnitude of the resistive dissipation predicted above. We can express it in terms of the time constant for the decay of the diamagnetic state of the superconductor, which we have already calculated in Sec. I. We have only to use for

[6] Considerations of this sort have been discussed by V. L. Ginzburg in his book *Superconductivity* (in Russian), Moscow (1946), and in Uspekhi Fiz. Nauk **42**, 169 (1950), and by P. M. Marcus at the Schenectady Cryogenics Conference, October, 1952 (unpublished). Particularly the latter anticipates the argument to be presented here.

[7] Dr. Herring has kindly pointed out to me that the formal expression of the assumption of thermal but not mechanical equilibrium is not immediately obvious in the presence of a magnetic field. The form that we have chosen is consistent with a treatment of the problem by the methods of irreversible thermodynamics, and is, in fact, just analogous to the corresponding treatments of the liquid helium problem.

[8] This is not at all unfamiliar. In an ordinary metal that has overlapping bands, so that there are two distinct groups of electrons, there is an internal circulation when a current flows in the presence of a magnetic field. The same applies to the holes and electrons in a semiconductor.

the impedance Z which loads the Hall voltage, some estimate of the resistance of the skin layer to the flow of normal currents. For an area of about 1 cm², a penetration defth of about 10^{-5} cm, and a resistivity of about $10^{-6}\Omega$ cm, the impedance may be estimated to be of the order of 10^{-11} ohm. This is more likely high than low for a typical sample. Thus the decay time, which was 5×10^{14} sec with a loading of one ohm, is reduced to a few hours at most. This is much too short, so that we can be sure that loading of the Hall voltage by the normal electrons does not occur on anything like the scale suggested by the two fluid model.

It is to be emphasized that this conclusion is independent of the possible existence of a contact potential that may cancel the Hall voltage at the surface of the specimen, and is a direct consequence of the systematic application of the two-fluid model to the problem. We can therefore only conclude that, if cancellation of the magnetic forces on the superelectrons takes place in the interior of the sample, then the forces responsible for this cancellation (the Hall forces) cannot act also on the normal electrons. This is a new feature of a superconductor, and is not contained in the existing phenomenological theories.

It is not hard to see that the new phenomenological theories[9] do not improve the situation, at least in their present form. What they provide for is a reluctance (because of the uncertainty principle) of the electrons to change their wave functions over distances shorter than $\sim10^{-4}$ cm, and this is not required by our considerations.

[9] Ginzburg and Landau, J. Exptl. Theoret. Phys. (U.S.S.R.) **20**, No. 12 (1950). J. Bardeen, Phys. Rev. **94**, 554 (1954).

IV. CONCLUSIONS

We have considered all possible explanations of the absence of a Hall effect in a superconductor, which both lie in the framework of existing theory, and are known to the author. All have been found wanting. The simple assumption of a contact potential that just cancels the Hall voltage has been found to beg the question, since we are still in trouble in the interior of the sample, as discussed in Sec. III.

We therefore conclude that an explanation of the absence of a Hall effect in a superconductor will require an extension of present theory in a direction not easily foreseen. It will require a long-range cooperative interaction among the electrons, that is not contained in the present phenomenological theories, and which will serve to transfer the magnetic forces on the superconducting electrons directly to the metallic lattice in such a way that the noncurrent carrying electrons are not affected. Whether the present beginnings of a molecular theory of superconductivity contain the seeds of such an interaction, we do not know.

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Theory of the Magnetic Susceptibility of Graphite*

JOHN E. HOVE
Nuclear Engineering and Manufacturing, North American Aviation, Inc., Downey, California

(Received March 4, 1955)

The theory of the magnetic susceptibility of graphite is presented in terms of a three-dimensional Wallace electron energy band structure. The experimentally observed variation with temperature is explained in a satisfactory manner, provided the interplane resonance integral occurring in the band approximation is given a value of about 0.5 ev. This is about five times larger than the previously used estimate and implies that a two-dimensional band approximation may be invalid in many cases. The in-plane resonance integral is obtained by fitting the variation, with electron concentration, of the electrical resistivity of a graphite-bisulphate residue compound. In this way a value of 1.63 ev for this integral is obtained. It might be noted that these values enabled a better fit of the resistivity over the entire range of bisulphatization than could be obtained by a two-dimensional theory. On the other hand, the value thus obtained for the actual magnitude of the susceptibility is lower than that observed by a factor of about 40. The (room temperature) variation of the susceptibility of bromine graphite is then analyzed on the basis of the above theory, using the indicated values of the constants. In this way, a value is obtained for the percentage of the bromine which is ionized. This is found to be weakly dependent on the amount of bromine, varying between 18% at 0.3 atomic percent bromine to 13% at 0.8 atomic percent bromine. The experimental value has been found to vary slightly around 18%. This agreement is very good and indicates that the theory is valid in explaining relative variations of the susceptibility, even though there is difficulty in predicting the absolute magnitude. The latter is the only serious discrepancy found in the present work and has not yet been explained.

I. INTRODUCTION

GRAPHITE shows a very high diamagnetic susceptibility which is, in addition, extremely anisotropic.[1] These properties have been explained by Eatherly[2] and Smoluchowski[3] as due to the highly anisotropic Brillouin zone structure of the conduction band.[4] It was originally shown by Peierls,[5] and later

* This paper is based on studies conducted for the U. S. Atomic Energy Commission.

[1] N. Ganguli and K. S. Krishman, Proc. Roy. Soc. (London) **177**, 168 (1940).

[2] W. P. Eatherly, see comments in discussion following reference 3.

[3] R. Smoluchowski, Revs. Modern Phys. **25**, 178 (1953).

[4] P. R. Wallace, Phys. Rev. **71**, 622 (1947).

[5] R. Peierls, Z. Physik **80**, 763 (1933).

extended by Wilson,[6] that there are terms in the magnetic susceptibility of crystalline conductors which depend on the curvature of the energy surfaces near the Fermi level and which can become large (and negative), especially if the band is nearly filled or nearly empty. These effects arise from changes in the level density rather than spin interactions and, for free electrons, reduce to the well-known Landau diamagnetism.[7] The earlier applications of this to graphite[2,3] utilized a two-dimensional approximation to the band structure[4] and were able to explain, in a qualitative way, the rapid changes in the susceptibility with such factors as bromination.[8] However, difficulties arose in making quantitative comparisons with experiment, particularly those relating to temperature effects. For this reason, a more detailed study was made using a three-dimensional band structure; the present paper gives the results of this study. In general, these results have been successful in explaining relative changes of the susceptibility. In addition, it has been possible to evaluate some of the constants inherent in the band structure by correlation of the theoretical and experimental susceptibility. The values of these constants thus obtained are of the order of magnitude of previous estimates (all of which are very approximate), but differ sufficiently to have a significant effect on the theoretical interpretation of the electrical properties of graphite.

There is one discrepancy between theory and experiment which the present work reduces, but does not remove. This concerns the absolute magnitude of the susceptibility. Previous attempts[9] to calculate this yielded a value which was lower than the observed by a factor of several hundred. The present calculation is still low by a factor of about forty. While this may be merely an error, either in the derivation or in the numerical work, all attempts to find such an error have failed. It should be emphasized that the theory of the magnetic susceptibility of electronic conductors presents some of the most difficult problems in solid-state physics and, at present, cannot be considered as generally satisfactory. This is especially true when the bands are degenerate (or nearly degenerate), as in the case of graphite, in which case terms which are usually considered negligible may become important.[6] Adams[9] has made some calculations for bismuth, where he shows that such terms are important, although it is very difficult to use his results for graphite. As will be seen, the use of the simpler (single band) theory of Peierls[5] for graphite is apparently satisfactory in explaining all relative variations of the susceptibility and fails only in predicting the magnitude. This would seem to imply that, if the trouble lies in neglecting terms of possible importance, these terms can be put into the form of the Peierls expression times a slowly varying function. The latter would then approximately cancel for the relative susceptibility.

[6] A. H. Wilson, Proc. Cambridge Phil. Soc. **49**, 292 (1953).
[7] L. Landau, Z. Physik **64**, 629 (1930).
[8] G. Hennig, J. Chem. Phys. **19**, 922 (1951); J. Chem. Phys. **20**, 1438, 1443 (1952).
[9] E. N. Adams II, Phys. Rev. **89**, 633 (1953).

II. TEMPERATURE DEPENDENCE

Using the nondegenerate theory of Wilson,[6] the conduction electron contribution to the magnetic susceptibility per unit volume is

$$\chi=\frac{\mu_0^2}{12\pi^2}\left(\frac{m}{\hbar^2}\right)^2\int\int\left[\frac{\partial^2E}{\partial k_x^2}\frac{\partial^2E}{\partial k_y^2}-\left(\frac{\partial^2E}{\partial k_x\partial k_y}\right)^2\right]\frac{\partial f_0}{\partial E}d^3\mathbf{k}, \quad (1)$$

for a constant magnetic field in the Z-direction. The integration is over all $\mathbf{k}$ space, f_0 is the Fermi-Dirac distribution function, and μ_0 is the Bohr magneton. The Wallace[4] model for the conduction band structure of graphite gives two touching bands, the lower one completely filled (at zero temperature) and the upper completely empty. If the origin of $\mathbf{k}$ is translated to one of the corners of the hexagonal Brillouin zone,[1] the Wallace energy values (near the corner) are

$$\epsilon=-\gamma_1\cos\frac{c}{2}k_z\pm\left[\gamma_1^2\cos^2\frac{c}{2}k_z+\tfrac{3}{4}\gamma_0^2a^2\kappa_{xy}^2\right]^{\frac{1}{2}}, \quad (2)$$

where ϵ is measured from the energy of band touching and the $(\pm)$ refers to the upper (empty) and lower bands. Here $\kappa_{xy}^2=\kappa_x^2+\kappa_y^2$ where $\kappa_x=k_x-k_y$ (corner), etc., and a and c are the magnitudes of the lattice translation vectors. The constants γ_0 and γ_1 are resonance integrals between coplanar and interplane nearest neighbors, respectively, and have been roughly estimated by Coulson[10] as being 0.9 ev and 0.09 ev. For the purposes of the present work, these will be regarded as disposable constants to be determined by experiment.

Using Eq. (2), it is a straightforward, although laborious, procedure to reduce Eq. (1) to the form

$$\chi=\frac{\mu_0^2}{\pi^2}\left(\frac{m}{\hbar^2}\right)^2\frac{a^2\gamma_0^2}{c\gamma_1(4kT)}\int_{-\infty}^{\infty}\operatorname{sech}^2\left(\frac{\epsilon-\Delta\zeta}{2kT}\right)h\left(\frac{\epsilon}{\gamma_1}\right)d\epsilon, \quad (3)$$

where $\Delta\zeta$ is the Fermi energy measured from the top of the filled band. The function $h(x)$ is symmetrical and has the following form for $x>0$:

$$h(x)=\frac{(2+x^2)}{2(1-x^2)^{5/2}}\left\{\ln\left[\frac{x}{2-x^2-[(1-x^2)(4-x^2)]^{1/2}}\right]-[(1-x^2)(4-x^2)]^{1/2}\right\}\quad\text{for } 0<x<1$$

$$=\frac{(2+x^2)}{2(x^2-1)^{5/2}}\left\{\frac{\pi}{2}-\tan^{-1}\left[\frac{(2-x^2)}{[(x^2-1)(4-x^2)]^{1/2}}\right]-[(x^2-1)(4-x^2)]^{1/2}\right\}\quad\text{for } 1<x<2$$

$$=\frac{\pi(x^2+2)}{2(x^2-1)^{5/2}}\quad\text{for } x>2.$$

[10] C. A. Coulson and R. Taylor, Proc. Phys. Soc. (London) **A65**, 815 (1952).

Thus for small x, $h(x)$ diverges logarithmically, while it has a slight peak at $x=2$ and approaches $\frac{1}{2}\pi x^{-3}$ for large x. The expansion of $h(x)$ for small x is

$$h(x)\simeq -\ln x - 0.614 + O(x^2);$$

it is valid to introduce this into Eq. (3) if $\gamma_1 > 4kT$. At room temperature, $4kT=0.1$ ev and the *a priori* assumption will be made that γ_1 is greater than this. Note that the approximation becomes better for lower temperatures.

For an untreated graphite single crystal, the quantity $\Delta\zeta$ is zero. Under this condition, and evaluating the constants of Eq. (3), there is found

$$-\chi_0 \simeq (0.375\times 10^{-7})\frac{\gamma_0^2}{\gamma_1}\left(\ln\frac{\gamma_1}{2kT}+0.27\right) \qquad (4)$$

in cgs mass units (ergs g^{-1} gauss^{-2}). This is correct up to an order of $(kT/\gamma_1)^2$. The variation of χ_0 with temperature has been measured by Ganguli and Krishnan,[1] Goldsmith,[11] McClelland,[12] and Owen.[13] In using these data, the core diamagnetism of the carbon atoms must be subtracted in order to obtain a value corresponding to the susceptibility considered here. The corrected experimental results are plotted in Fig. 1, normalized to 300°K. Superimposed on this plot are curves of Eq. (4) (normalized similarly) for several values of γ_1. It is apparent from this that the best fit is obtained when γ_1 is between 0.5 and 0.8 ev, with 0.5 ev appearing preferable. The commonly used value of 0.1 ev gives worse agreement with the observed results than can be accounted for on the basis that one is approaching the limit of validity inherent in Eq. (4) at room temperature. Thus, in the work to follow, a value of 0.5 ev will be used for γ_1. If χ_0 is plotted against $\ln T$, a straight line results at low temperatures (less than about room temperature) with a slope proportional to $(\gamma_0^2\gamma_1^{-1})$. This slope, with a γ_1 of 0.5 ev, gives a value of 10.5 ev for γ_0. The same value of γ_0 is obtained from fitting the experimental magnitude of the susceptibility (about 22.1×10^{-6} cgs units). Previously estimated values[10,14] of γ_0 lie between 1.0 and 3.0 ev and it is not likely that the order of magnitude of these former values is wrong. Furthermore, as will be shown later, a γ_0 of 10 ev gives serious discrepancies with the brom-graphite data, which require a value of about 1.6 ev. Thus, as mentioned in the previous section, it must be concluded that the coefficient in Eq. (1) is low by a factor of about forty. At present this cannot be explained. Since all of the relative changes of χ appear to be explained by Eq. (1), however, it is reasonable to assume that whatever causes this discrepancy does not greatly alter the form of the energy integral; similarly, it probably does not depend on a serious alteration of the assumed energy band structure [Eq. (2)].

[11] M. Goldsmith, J. Chem. Phys. **18**, 523 (1950).
[12] J. D. McClelland (private communication).
[13] M. Owen, Ann. Physik **37**, 657 (1912).
[14] K. Komatsu and T. Nagamiya, J. Phys. Soc. Japan **6**, 438 (1951).

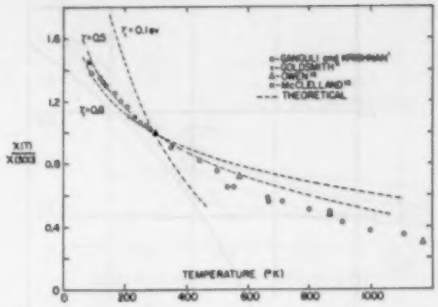


FIG. 1. The relative variation with temperature of the conduction electron diamagnetic susceptibility of normal graphite.

III. BROM-GRAPHITE RESIDUE COMPOUNDS

The electronic properties of many of the interstitial compounds of graphite have been extensively studied by Hennig.[8] The compound of particular interest to the present report is that with bromine. The brom-graphite residue compounds are useful in studying the properties of graphite because a fraction of the bromine is always ionized, which thus removes electrons from the graphite band. Since the bromine is presumably at crystal boundaries,[8] it has little effect on the band structure itself. The fraction of the residual bromine which is ionized is fairly constant (at about 18%) for ion concentrations of from 0.01 to 0.09 atomic percent. This ionized fraction was obtained by Hennig by comparing the electrical resistivity with that for graphite-bisulphate compounds, for which the ion concentration can be found with a reasonable accuracy by electrolytic methods. Since the ion concentration also represents the number of electrons which are removed from the band, the brom-graphite residue compounds permit a study of various electronic properties of graphite as a function of various known electron concentrations. In the present paper, the experimental susceptibility of brom-graphite, measured by Hennig and McClelland,[15] will be compared with the theory.

The effect of depleting electrons from the graphite conduction band will be to alter $\Delta\zeta$, the Fermi energy measured from the top of the band, in Eq. (3) for the magnetic susceptibility. Although $\Delta\zeta$ will actually be negative, χ is a symmetrical function of $\Delta\zeta$ [because of the use of Eq. (2) for the energy spectrum] and $\Delta\zeta$ may be taken as a positive quantity only. If $\Delta\zeta$ is small (considerably less than γ_1), the function $h(\epsilon/\gamma_1)$ in the integrand may be expanded as was done in the last section. Then the susceptibility (in cgs mass units) becomes

$$-\chi \simeq (0.375\times 10^{-7})\frac{\gamma_0^2}{\gamma_1}\left\{\ln\frac{\gamma_1}{2kT}-0.614 + \tfrac{1}{2}F\left(\frac{\Delta\zeta}{2kT}\right)\right\}, \qquad (5)$$

[15] G. Hennig and J. D. McClelland, J. Chem. Phys. (to be published).

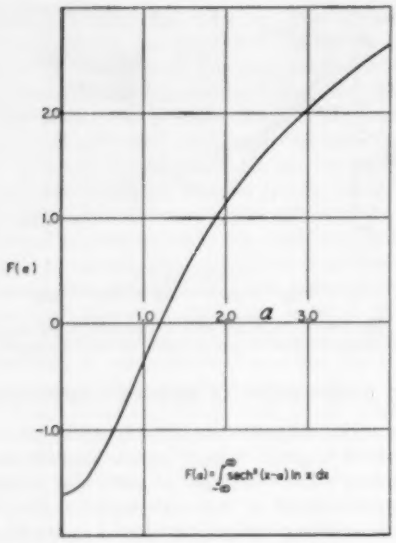


FIG. 2. A plot of the function $F(\alpha)$.

where

$$F(\alpha)=-\int_{-\infty}^{\infty}\operatorname{sech}^2(x-\alpha)\ln x dx.$$

The function F must be found by numerical integration but, once obtained, it is known for any temperature. It is given in Fig. 2 as a function of its argument.

If $\Delta\zeta$ is much larger than $4kT$, the integrand in Eq. (3) has an appreciable value only at $\epsilon=0$ and $\epsilon=\Delta\zeta$. Then the susceptibility is

$$-\chi\simeq(0.375\times10^{-7})\frac{\gamma_0^2}{\gamma}\times\left\{\frac{\gamma_1}{32kT}\exp\left[-\tfrac{1}{4}\exp\frac{\Delta\zeta}{kT}\right]+h\left(\frac{\Delta\zeta}{\gamma_1}\right)\right\}. \quad (6)$$

The first term corresponds to the integration around $\epsilon=0$ and is negligible. The remaining term is temperature independent and is the analog of the Landau susceptibility of a degenerate free electron gas. It is, of course, evident that if enough electrons are trapped out of the energy band, graphite must look nearly metallic.

Since, from the last section, γ_1 has a value of about 0.5 ev, the ranges of validity of Eqs. (5) and (6) overlap at low temperatures and nearly do so even at room temperature. Thus, a satisfactory picture of Eq. (3) may be obtained by plotting Eqs. (5) and (6) and connecting them smoothly. This is done in Fig. 3, which actually plots χ/χ_0, where χ_0 is the value for $\Delta\zeta=0$. Two temperatures, 78° and 288°K, are shown. The experimental points of Hennig and McClelland[15] for type AGOT-KC graphite are given in Fig. 4 as a function of q, the atomic percent of bromine in the residue compound. A cross plot of Figs. 3 and 4 yields the variation of $\Delta\zeta$ with bromine concentration, which is shown in Fig. 5. As can be seen from Fig. 5, one may very nearly put

$$\Delta\zeta=0.054q^{\frac{1}{3}}, \quad (7)$$

where q is atomic percent of bromine and $\Delta\zeta$ is in ev.

It is necessary to know the relationship between the number of electrons trapped (n_e) and $\Delta\zeta$. First of all, it can be shown that, at room temperature and below, the temperature dependence of $\Delta\zeta$ is negligible for graphite (the variation is not more than about 10%). Then, one can relate

$$n_e=\int_0^{\Delta\zeta}N(\epsilon)d\epsilon,$$

where $N(\epsilon)$ is the density of states per carbon atom if n_e is the number of trapped electrons per atom. Using the energy spectrum of Eq. (2), it is easily shown that

$$n_e\simeq\frac{2\gamma_1(\Delta\zeta)}{\sqrt{3}\pi^2\gamma_0^2} \quad \text{for} \quad \Delta\zeta\ll\frac{4}{\pi}\gamma_1. \quad (8)$$

For a γ_1 of 0.5 ev, this is a valid approximation insofar as the present work is concerned. The value of γ_0 can be independently obtained from the variation of the electrical resistivity with the residue graphite-bisulphate compound. This has also been studied by Hennig,[8] and, by electrolytic methods, the number of trapped elec-

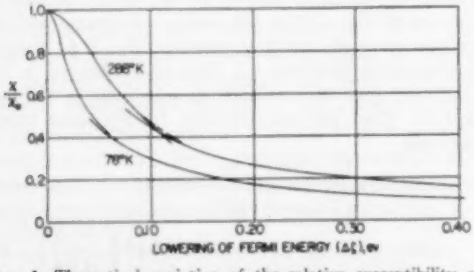


FIG. 3. Theoretical variation of the relative susceptibility of graphite with Fermi energy [from Eqs. (5) and (6)].

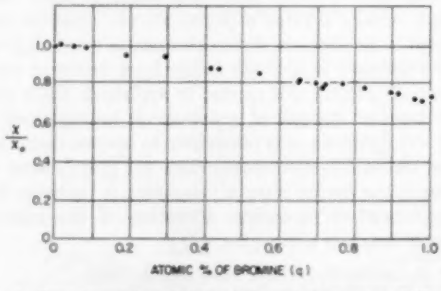


FIG. 4. Experimental variation of the relative susceptibility of brom-graphite with amount of bromine (from reference 15).

trons has been directly determined. Using the band structure of Eq. (2), the resistivity, normalized to the untreated resistivity, can be written[8]

$$\frac{\rho}{\rho_0}=(\ln 2)\Big/\ln\left(2\cosh\frac{\Delta\zeta}{2kT}\right). \tag{9}$$

It has been assumed here that $\Delta\zeta$ is considerably less than γ_1 and that the electron relaxation time is energy independent; the latter would be expected if boundary or defect scattering were predominant. Using Eq. (8), Eq. (9) may be written in terms of the number of electrons trapped per atom, with γ_0 as a parameter. On Fig. 6 is reproduced the resistivity variation found by Hennig[8] for room temperature and 144°K, together with a plot of Eq. (9) for $\gamma_1=0.5$ ev and $\gamma_0=1.63$ ev. The room temperature fit is good but it becomes worse for the low temperature. The latter may be because the scattering is not strictly energy independent or that the band structure is not sufficiently accurate. It should be mentioned that as long as γ_1 is much larger than both $\Delta\zeta$ and kT, the theoretical two- and three-dimensional resistivities are identical in terms of $\Delta\zeta$, but a two-dimensional energy band model gives n_e proportional to $(\Delta\zeta)^2$. A two-dimensional fit[8] is also shown in Fig. 6, and it can be seen that this is less satisfactory than the three-dimensional result.

Using the values 0.5 ev and 1.63 ev thus obtained for γ_1 and γ_0, the relation between the number of trapped electrons (per C atom) and the Fermi energy lowering (in ev) becomes

$$n_e\simeq 2.2\times 10^{-2}\Delta\zeta. \tag{10}$$

If p is the atomic percent of the bromine which is ionized, then n_e may be written as $(10^{-4}p)q$. Introducing Eq. (10) into Eq. (7) (the relation between $\Delta\zeta$ and q as found from the susceptibility), there is found

$$p\simeq 12q^{-\frac{1}{3}}\%. \tag{11}$$

Thus, p varies from 18% ionization for $q=0.3$ atomic percent of bromine to 13% ionization at $q=0.8$ atomic percent of bromine. This is in good agreement with the estimates which Hennig obtained independently (which varied slightly around a value of 18%). It is especially gratifying to note that the values found from the magnetic susceptibility data involve no arbitrary parameters, i.e., the constants γ_1 and γ_0 were obtained by completely independent methods.

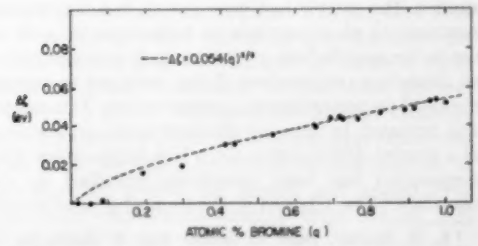


FIG. 5. Variation of the Fermi energy ($\Delta\zeta$) of brom-graphite with amount of bromine as determined from the susceptibility variation.

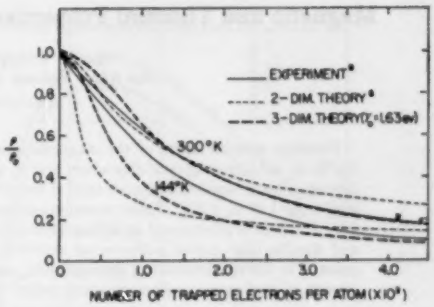


FIG. 6. Relative variation of the electrical resistivity of brom-graphite with electron concentration.

IV. CONCLUSIONS

The major conclusion arising from the present work is that the relative changes of the magnetic susceptibility of graphite with temperature and with electron concentration can be quantitatively explained by a simple single-band theory if a three-dimensional energy band spectrum is used. This requires an in-plane resonance integral of 1.63 ev and an interplane integral of about 0.5 ev. The latter is considerably higher than the estimate made by Coulson[10] and this implies that a two-dimensional graphite model may be an invalid approximation for many of the electronic properties. It does not, however, appear that this simple theory can predict the actual magnitude of the observed susceptibility. It is not yet clear whether this discrepancy (a factor of about forty) can be removed by a multiband theory[9] or whether it is due to a serious fault in the graphite band structure. For example, Johnston[10] has pointed out that there is a slight band overlap inside the zone; it is conceivable that this could contribute a sizeable diamagnetism, although it would seem highly fortuitous if this effect showed the same relative behavior as found in the present paper. Because of the success of the present treatment of the relative variation, however, it is a reasonable presumption that any neglected effects of importance can be put into the form of a term like Eq. (1) multiplied by a slowly varying function of temperature and electron concentration. The latter term would then approximately cancel in a relative expression.

ACKNOWLEDGMENTS

The author is indebted to W. P. Eatherly, J. A. Krumhansl, and G. W. Lehman for discussions of this problem.

[10] D. S. Johnston, Atomic Energy Research Establishment, Harwell Report AERE-855, 1952 (unpublished).

Magnetic and Thermal Properties of UI_3 at Liquid Helium Temperatures*

L. D. Roberts and R. B. Murray†
Oak Ridge National Laboratory, Oak Ridge, Tennessee
(Received June 22, 1955)

Previous measurements of the magnetic susceptibility of UI_3 revealed a susceptibility maximum near 3.2°K of antiferromagnetic character, and a second susceptibility maximum at 1.5°K which is strongly depressed by an applied magnetic field of only 5 to 10 oersteds. The specific heat of UI_3 has now been measured from 1.2° to 4.2°K. These measurements show a λ-type anomaly of magnitude 10 joules/mole-deg at 2.61°K which is interpreted as arising from the antiferromagnetic transition. The behavior of the entropy and specific heat curves indicates an extensive short-range order above the antiferromagnetic transition point. The lower temperature susceptibility anomaly is tentatively interpreted as a manifestation of the growth of long-range antiferromagnetic order.

SUSCEPTIBILITY MEASUREMENTS

THE magnetic susceptibilities of UCl_3, UBr_3, and UI_3 have recently been measured from 77° to ~500°K by Dawson.[1] In the region near room temperature and above the susceptibilities of these compounds followed the Curie-Weiss law. In the case of UI_3, the susceptibility followed a Curie-Weiss law in the temperature region 200°–400°K, with a measured Weiss constant of about 5°K; this small Weiss constant suggests the possibility of a magnetic ordering transition at liquid helium temperatures.

In order to investigate the magnetic properties of UI_3 at very low temperatures, Roberts, Lavalle, and Erickson[2] measured the differential magnetic susceptibility of polycrystalline UI_3 in the region 1.1°–4.2°K. These measurements were carried out with a mutual inductance bridge which has been described previously.[3] The susceptibility curves are reproduced here in Figs. 1 and 2 along with some additional new data. These data were obtained as before[2] with an oscillating measuring field of about 2 oersteds amplitude and a frequency of 500 cycles per second. The temperature of the sample was obtained from liquid helium vapor pressure mealsurements using the 1948 temperature scale[4] with corrections given by Erickson and Roberts.[5] The bridge calbration was determined at 77°K using the absolute susceptibility as measured by Dawson at that temperature. The results of our measurements show a well-defined susceptibility maximum at 3.2°K, characteristic of a transition to the antiferromagnetic state. In addition to the 3.2°K maximum, the susceptibility displays further anomalous behavior in that both the reactive component (χ') and the resistive component (χ'') pass through a maximum near 1.5°K. The 1.5°K maximum in both components of the susceptibility is strongly depressed by applied magnetic fields of 5–10 oersteds, while the 3.2°K maximum is relatively field independent.

The open and closed circles in Figs. 1 and 2 represent data points taken on two different samples of UI_3 which were prepared by different chemical processes[2] from materials of high purity. The very good agreement between these two samples is taken as evidence that this anomalous susceptibility behavior is an inherent property of UI_3, and suggests, in particular, that the exceptional behavior near 1.5°K is not due to impurity effects. It is interesting to note that the lower temperature susceptibility maximum would not have been observed if the usual strong magnetic field Gouy method had been used rather than the above bridge procedure.

SPECIFIC HEAT MEASUREMENTS

In order to extend the study of the properties of UI_3 at liquid helium temperatures, the specific heat of polycrystalline UI_3 has been measured from 1.2°–4.2°K by the usual techniques of low temperature calorimetry. The over-all procedures and apparatus used will be described elsewhere.[6] The temperature scale used in the specific heat measurements is that noted in reference 5. The sample of UI_3 used in the specific heat measurements (0.1041 mole) was part of a large sample prepared at the time of the susceptibility studies. The sample had been stored in a sealed glass container in an atmosphere of helium gas prior to its use in the specific heat experiments. It was handled at all times in an atmosphere of dry, inert gas to prevent chemical decomposition or contamination. This sample was prepared by reacting uranium metal, in the form of a powder, with iodine gas at a high temperature. This preparation has been previously described in the literature.[7] The UI_3 was in the form of a powder with

* Reported at the New York meeting of the American Physical Society, January, 1955.

† Graduate Fellow of the Oak Ridge Institute of Nuclear Studies from the University of Tennessee. This work was included in a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Tennessee.

[1] J. K. Dawson, Atomic Energy Research Establishment, Report C/R 578, September, 1950 (unpublished). See also J. Chem. Soc., 429 (1951).

[2] Roberts, Lavalle, and Erickson, J. Chem. Phys. **22**, 1145 (1954).

[3] Erickson, Roberts, and Dabbs, Rev. Sci. Instr. **25**, 1178 (1954).

[4] H. Van Dijk and D. Shoenberg, Nature **164**, 151 (1949).

[5] R. A. Erickson and L. D. Roberts, Phys. Rev. **93**, 957 (1954). These corrections are in accord with the temperature scale adopted at the Paris Low Temperature Conference, September, 1955.

[6] R. B. Murray, paper on specific heat of $MnCl_2$ (to be published).

[7] J. J. Katz and E. Rabinowitch, *The Chemistry of Uranium* (McGraw-Hill Book Company, New York, 1951).

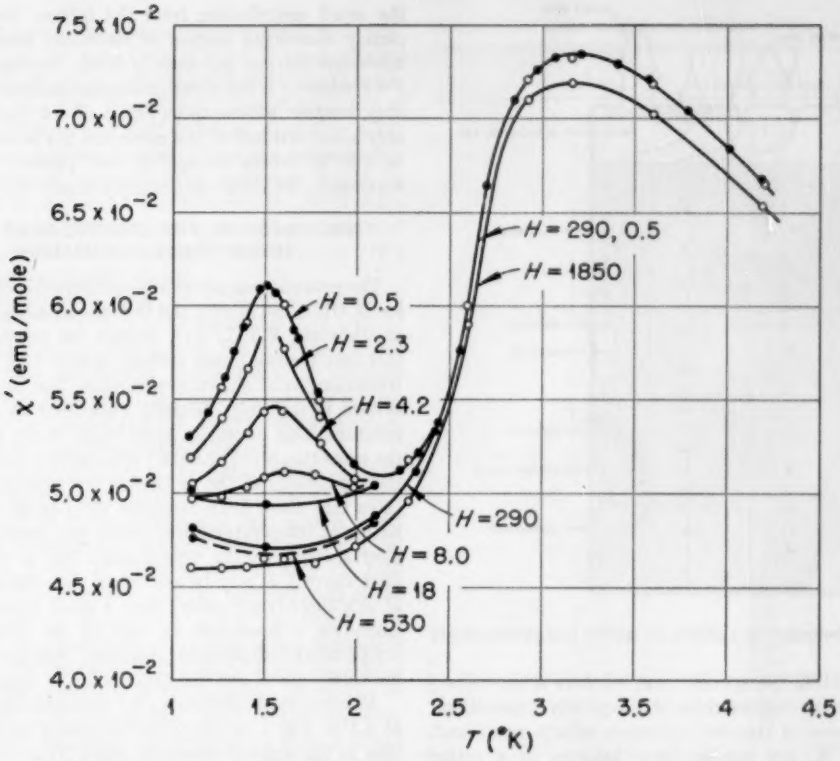


FIG. 1. Reactive component of the differential magnetic susceptibility of polycrystalline UI_3 as a function of temperature and applied magnetic field. The applied field H was parallel to the oscillating measuring field. H is in oersteds.

average particle size $\sim\frac{1}{4}$ mm. The UI_3 samples used in the susceptibility and specific heat studies were prepared by D. E. Lavalle of this laboratory.

The copper calorimeter cup used in the UI_3 experiments is shown schematically in Fig. 3. This cup was gold-plated on the inside to prevent any chemical reaction between the sample and copper wall. A thin-wall, gold-plated, copper sleeve was bonded to the bottom of the calorimeter cup with Araldite cement. A heater coil of Constantan wire was insulated from and bound to the outside of the sleeve with Araldite. The carbon resistance thermometer was prepared from a $\frac{1}{2}$-watt Allen-Bradley resistor by grinding away the external insulating cover. This unit (47 ohms at room temperature) was cemented to the wall of the cup and was covered by the UI_3 specimen. Helium exchange gas at a pressure of 20 cm Hg at room temperature was sealed inside the cup to insure thermal equilibrium. In the measuring process the sample was heated over temperature increments ranging from 20 to 100 millidegrees.

The results of the specific heat measurements are shown in Fig. 4 and in Table I. The data points given here represent the molar specific heat of UI_3, the specific heat of the calorimeter cup, thermometer, etc., having been subtracted from the measured points. This correction, which amounted at most to only 2% of the total specific heat at 4.2°K, was calculated from known specific heats of the components of the calorimeter assembly.

The specific heat maximum at 2.61°K (Fig. 4) is interpreted as arising from the onset of long-range antiferromagnetic order, and is thus to be associated with the susceptibility maximum observed at 3.2°K (Fig. 1).

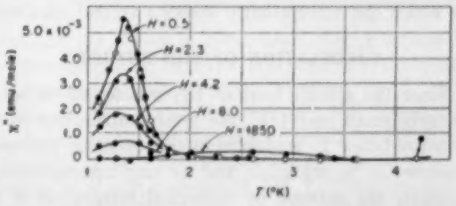


FIG. 2. Resistive component of the differential magnetic susceptibility of polycrystalline UI_3 as a function of temperature and applied magnetic field. The applied field H was parallel to the oscillating measuring field. H is in oersteds.

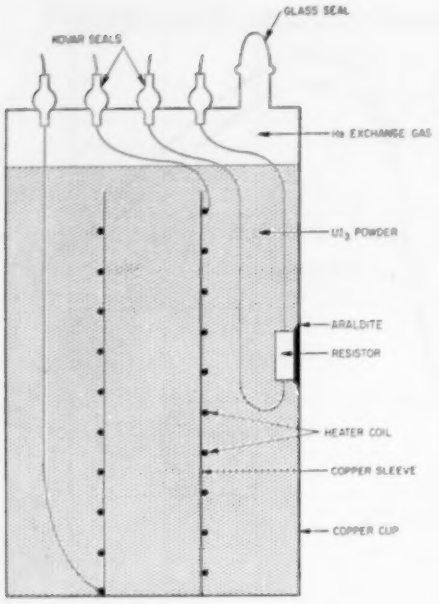


FIG. 3. Calorimeter cup assembly for specific heat measurements

Below 2.61°K the specific heat exhibits a very sharp fall which is characteristic of cooperative transitions. No evidence of thermal hysteresis effects was found. Above 3°K, the specific heat behaves in a rather unusual manner, decreasing very slowly with increasing temperature. As no specific heat data are available at temperatures above 4.2°K, it is not possible to establish with any accuracy the lattice contribution to the specific heat in the liquid helium region. By comparison with the lattice specific heat of other compounds at liquid helium temperatures, however, we may estimate the order of magnitude of the lattice specific heat to be ~0.1 to 0.2 joule/mole-deg at 4°K. With this estimate the lattice specific heat amounts to about 3 to 6% of the total specific heat at 4°K. Thus it does not seem reasonable to account for the large specific heat above 3° as arising from the lattice contribution.

No indication of a second specific heat anomaly was found in the neighborhood of 1.5°K, the temperature at which the susceptibility shows a second maximum.

CALCULATION OF THE ENTROPY

Since the specific heat of UI_3 was found to be extremely small near 1°K, the total contribution to the entropy below 1° arising from this ordering process is assumed to be negligible, and we may correspondingly evaluate the entropy by numerical integration of the specific heat data between 1.2° and 4.2°K. The results of the entropy calculation are shown in Fig. 5. The entropy calculated in this manner includes, of course, the small contribution from the lattice. For a completely disordered system of electronic moments the minimum entropy per mole is $R \ln 2$, corresponding to $J=\frac{1}{2}$, where J is the effective angular momentum quantum number. Reference to Fig. 5 shows that at 4.2°K only about one half of this minimum has been achieved. At 2.61°K, where the specific heat passes through its maximum, the calculated entropy is only 0.2 $R \ln 2$.

DISCUSSION OF THE SPECIFIC HEAT AND SUSCEPTIBILITY ANOMALIES

The ground-state electronic configuration of the U^{+++} ion in UI_3 is not known and thus the maximum entropy of disorder, $R \ln(2J+1)$, cannot be predicted. The fact that the calculated entropy is only $0.2R \ln 2$ at the transition point shows nevertheless that the electron system is still only partially disordered at that temperature. This unusually small value of the entropy at the transition point (2.61°K) is in distinct contrast with the predictions of the molecular field theory and indicates that there is an extensive short-range order persisting at temperatures well above the transition point. Correspondingly, the very gradual fall in the specific heat above 3°K may be associated with the slow decay of this short-range order. Such a small entropy at the transition temperature as well as the large, almost temperature-independent magnetic specific heat immediately above the transition point, is unusual.

As mentioned previously, the susceptibility anomaly at 3.2°K, Fig. 1, is interpreted as arising from a transition to the antiferromagnetic state. This conclusion is based on the fact that (1) the susceptibility demonstrates a well-defined maximum which is nearly field independent, and (2) the specific heat exhibits a λ-type transition characteristic of cooperative ordering. In view of the evidence of an extensive short-range order, it is not surprising that the peak in the specific heat occurs at a somewhat lower temperature than does the susceptibility maximum, since the specific heat anomaly presumably occurs with the onset of long-range magnetic ordering, whereas the susceptibility would be affected by the extensive short-range order above this transition.

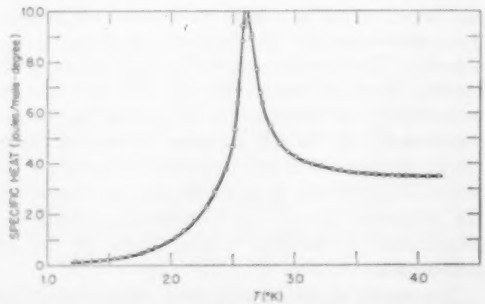


FIG. 4. Molar specific heat of UI_3 as a function of temperature.

Turning now to the susceptibility anomaly near 1.5°, we note that the magnitude of the 1.5° susceptibility peak is of the same order as the antiferromagnetic maximum at 3.2°K. The extreme field dependence of the susceptibility near 1.5° indicates that this effect is also of a cooperative nature, since the susceptibility of a system of noninteracting moments would be unaffected by such small applied fields. In addition, the fact that the susceptibility is depressed by these small fields suggests that the ordering is, at least in part, of a weakly "ferromagnetic" nature.

The presence of a large out-of-phase susceptibility component, χ'', Fig. 2, is to be associated with a phase lag between the applied measuring field and the induced moment. This behavior is characteristic of ordinary ferromagnetic materials (e.g., iron) suggesting further an analogy between the 1.5°K susceptibility anomaly and a type of "ferromagnetic" behavior.

The extreme sensitivity of the susceptibility at 1.5°K to small magnetic fields has been noted above. We now observe that, as the applied steady field attains larger values the susceptibility appears to approach a lower limit. This occurs at fields in the vicinity of 1850 oersteds. The susceptibility in the neighborhood of the antiferromagnetic maximum at 3.2°K was found to be essentially field independent; it then seems reasonable to ascribe the strong field limiting susceptibility near 1.5°K to the residual antiferromagnetic properties of the system. This strong field limit of the antiferromagnetic susceptibility extrapolated to 0°K is about 0.6 times its maximum value at 3.2°K. Thus there appears to be a correlation with the Van Vleck two-sublattice model of antiferromagnetism,[8] which predicts a ratio of $\frac{2}{3}$. It does not seem justifiable, however, to conclude on this basis alone that the antiferromagnetic order in UI_3 is of a simple two-sublattice type. Rather, we note the qualitative similarity in the behavior of the residual antiferromagnetic susceptibility observed here, and that predicted by the Van Vleck theory.

If we ascribe differences between the 1850 oersted curve and the curves in lower fields, Fig. 1, as due to the "ferromagnetic" behavior, it is possible to determine this incremental susceptibility, denoted by S, as a function of field. A plot of S *vs* H at 1.5°K is shown in Fig. 6. The corresponding magnetization (at 1.5°K) as a function of field is obtained by integrating S as a function of H and is shown in Fig. 7. This integration neglects the small (about one percent) contribution from the resistive component. By integrating this susceptibility to large fields it was found that the incremental moment appears to saturate at $M_{sat} \approx 0.6$

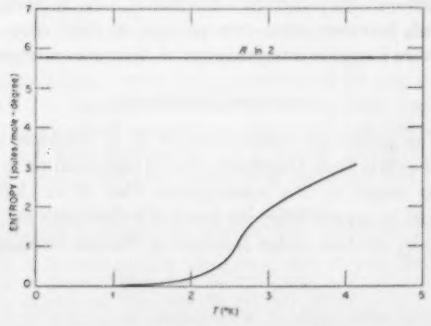


FIG. 5. Entropy of UI_3 as a function of temperature.

TABLE I. Molar specific heat of UI_3.

T (deg K)	C_{max} (joules/mole-deg)
Experiment Number 1	
1.230	0.173
1.287	0.129
1.355	0.175
1.424	0.172
1.490	0.212
1.547	0.250
1.598	0.293
1.663	0.359
1.766	0.496
1.820	0.575
1.875	0.679
1.963	0.882
2.042	1.11
2.105	1.33
2.191	1.72
2.291	2.30
2.363	2.89
2.439	3.72
2.500	4.71
2.537	6.39
Experiment Number 2	
2.443	3.78
2.518	5.35
2.579	8.84
2.648	8.96
2.721	6.81
2.803	5.47
2.883	4.77
2.949	4.45
3.015	4.26
3.083	4.07
3.144	3.97
3.229	3.84
3.306	3.78
3.377	3.71
3.446	3.64
3.517	3.60
3.589	3.58
3.659	3.55
3.724	3.52
3.790	3.53
3.846	3.50
3.904	3.50
3.992	3.48
4.079	3.46
4.164	3.48
2.496	4.65
2.548	6.90
2.576	9.44
2.599	9.43
2.622	9.66
2.651	9.09
2.691	7.71
2.737	6.40

[8] J. H. Van Vleck, J. Chem. Phys. **9**, 85 (1941).

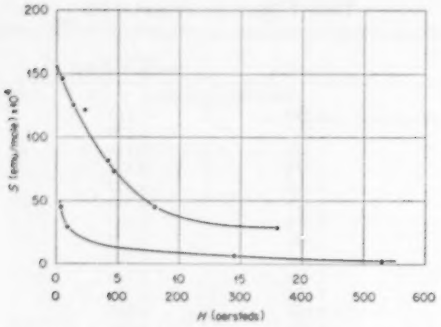


FIG. 6. "Ferromagnetic" susceptibility of UI_3 as a function of applied field H at 1.5°K.

emu/mole (at 1.5°K) or an average moment of 10^{-24} emu per U^{+++} ion. This is quite small, being of the order of 10^{-4} Bohr magneton. This ratio of 10^{-4} should be taken as an approximate value, since we have assumed that the residual antiferromagnetic susceptibility is field independent. This is not strictly true, as the susceptibility near 3°K is in fact depressed somewhat by fields greater than about 300 oersteds and a large contribution to the area under the S *vs* H curve, Fig. 6, comes at large fields. This calculation, nevertheless, should indicate the correct order of magnitude.

Because of the approximate nature of the above calculation, we do not present detailed results for M_{sat} as a function of temperature. We note, however, that the extreme field sensitivity of the susceptibility first appears at a temperature just below the temperature of maximum specific heat and displays a trend toward increasing magnetization with decreasing temperature. Thus, the extreme field sensitivity of the susceptibility and the growth of long range antiferromagnetic order are simultaneous phenomena in temperature and are possibly correlated. In this connection it is also interesting to note that with small applied fields the phase angle $\theta = \arctan\chi''/\chi'$ also increases with decreasing temperature below about 2.6°K.

Referring to the specific heat measurement, we note that there is only one specific heat anomaly, although there are two susceptibility maxima. This suggests either that both of the susceptibility peaks are associated with the same ordering process, or that one of the susceptibility maxima (presumably that at 1.5°) results from the ordering, perhaps by some new process, of relatively few moments.

Weak ferromagnetic properties are often found to accompany antiferromagnetic transitions. Néel[9] has suggested a possible explanation of this for crystals of a layer lattice structure, of which UI_3 is an example, by postulating the existence of antiferromagnetic domains. These antiferromagnetic domains are considered by Néel to consist of ferromagnetic sheets of alternating spin orientation with a corresponding zero net magnetization. The boundary between two antiferromagnetic domains is considered to be a region (a few layers in thickness) in which the spin phase shifts by 180 degrees, giving rise to an uncompensated net spin, of a ferromagnetic nature, in the separating wall. On the basis of Néel's picture, the growth of magnetization in the separating walls, and thus of the "ferromagnetic" effects, would be expected to be simply correlated with the growth of long-range antiferromagnetic order throughout the crystal. In the case of UI_3, the observed parasitic "ferromagnetism" begins to grow in at temperatures just below the 2.61°K antiferromagnetic specific heat anomaly as previously noted. Thus there is an interesting parallel between Néel's suggestion of ferromagnetic walls separating antiferromagnetic domains and the observed behavior of UI_3. It must be noted, however, that this picture of Néel does not provide a unique interpretation of these measurements.

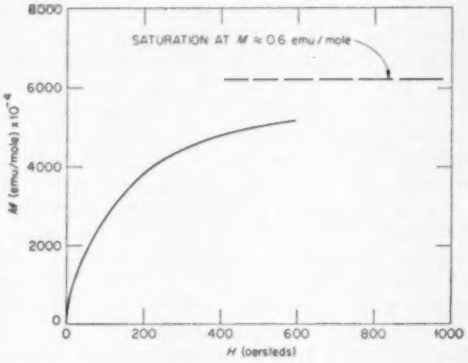


FIG. 7. "Ferromagnetic" moment of UI_3 as a function of applied field H at 1.5°K.

ACKNOWLEDGMENTS

The authors are indebted to Dr. R. A. Erickson, now at the Ohio State University, for his participation in the early stages of this investigation. One of us (RBM) wishes to acknowledge the grant of a Graduate Fellowship by the Oak Ridge Institute of Nuclear Studies.

[9] L. Néel, Brussels Conference, 1954 (unpublished).

Relative Absorption of 10 Mc/sec Longitudinal Sound Waves in a Superconducting Polycrystalline Tin Rod*

L. MACKINNON†
Physics Department, Brown University, Providence, Rhode Island
(Received June 6, 1955)

The relative absorption of 10 Mc/sec sound pulses has been studied between 1.5°K and 3.73°K (the transition temperature) in superconducting polycrystalline tin and above 3.73°K to about 4.2°K in the normal metal. Above the transition the absorption was heavy, but below it, the absorption decreased considerably as the temperature was lowered. It was possible to estimate how the minimum value of the absorption coefficient (which differed from the true value by an undetermined constant equal to the absorption coefficient at the absolute zero of temperature) varied as a function of temperature but no success has yet been achieved in correlating this quantitatively with existing theories of superconductivity or of absorption of sound in metals at low temperatures.

I. INTRODUCTION

THE transition between the superconducting and normal phases was first shown to have an effect on the absorption of sound in a pure superconductor by Bömmel and Olsen,[1] who observed that the absorption at a frequency of 1 Mc/sec in lead changed while the transition was taking place in the presence of a magnetic field; that the absorption was in fact different in the two phases for a lead single crystal was later discovered by Bömmel,[2] who covered the frequency range 9–27 Mc/sec and showed that in the superconducting phase the absorption was substantially less than in the normal phase for the same temperature both for longitudinal and transverse waves. That a similar effect was observable in polycrystalline tin with 10 Mc/sec (longitudinal) sound was reported by the author[3] and it is the purpose of this paper to amplify the contents of that abstract. A number of other investigators[4–6] have also recently reported an essentially similar effect in tin.

II. EXPERIMENTAL METHOD

The experimental arrangement is illustrated by the block diagram (Fig. 1). Short (1–2 μsec) pulses of high frequency (10 Mc/sec) alternating voltage were fed to an X-cut quartz transducer attached to one end of the specimen. There was sufficient electrical "pickup" from this pulse to produce a deflection (after amplification, etc.) on the cathode-ray oscilloscope and also to provide a convenient triggering and synchronizing impulse for the sweep (of approximately 50 μsec duration). The transducer converted the electrical impulse into longitudinal sound waves which traveled down the rod and were reconverted by a second similar transducer into an electrical impulse at the other end. The envelope of this transmitted pulse was visible (after amplification and rectification) on the sweep of the oscilloscope as shown in the diagram (and also in Figs. 3 and 4).

The tin specimen was made from rods of 99.999% pure tin (Johnson, Matthey and Company, Laboratory No. 5606, Catalogue No. JM 530), and was cast in vacuo after degassing as a rod of about 2.2 cm diameter. The ends were turned flat in a lathe and then ground with No. 600 grind. The quartz plates were of ½-in. diameter and unpolished. The bonding material between transducer and specimen was Dow Corning 200 fluid, viscosity 2.5×10^6 centistokes, as recommended by McSkimin[7] and the bond was made substantially as described by Squire[8] except that the electrodes probably pressed down rather harder on the specimen. The method of mounting the specimen is illustrated in Fig. 2; the tin rod was upright in the liquid helium bath, suspended by the electrical leads, and the cryostat was of conventional design. Temperatures were measured on the 1949 helium vapor pressure scale.

The electronic circuits were part of an apparatus described recently by Cedrone and Curran[9]; the oscilloscope was a Tektronix type 514D and the trace was photographically recorded with a Fairchild Polaroid Oscilloscope Camera. The amplifier was tested for

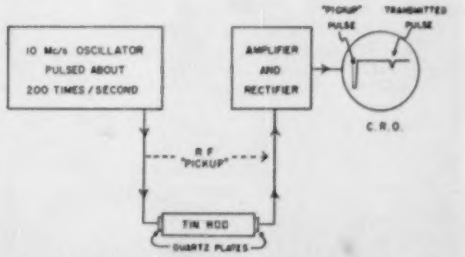


FIG. 1. Block diagram of the experimental arrangement.

* Supported by the Office of Naval Research.
† On leave of absence from Charterhouse, Godalming, Surrey, England.
[1] H. E. Bömmel and J. L. Olsen, Phys. Rev. **91**, 1017 (1953).
[2] H. E. Bömmel, Phys. Rev. **96**, 220 (1954).
[3] L. Mackinnon, Phys. Rev. **98**, 1181(A) (1955).
[4] H. E. Bömmel, as reported by W. P. Mason, Phys. Rev. **97**, 557 (1955).
[5] Benjamin Welber, Phys. Rev. **98**, 1196(A) (1955).
[6] J. K. Landauer, Phys. Rev. **96**, 296 (1954).
[7] H. J. McSkimin, J. Appl. Phys. **29**, 988 (1953).
[8] C. F. Squire, Phys. Rev. **95**, 1126 (1954).
[9] N. P. Cedrone and D. R. Curran, J. Acoust. Soc. Am. **26**, 963 (1954).

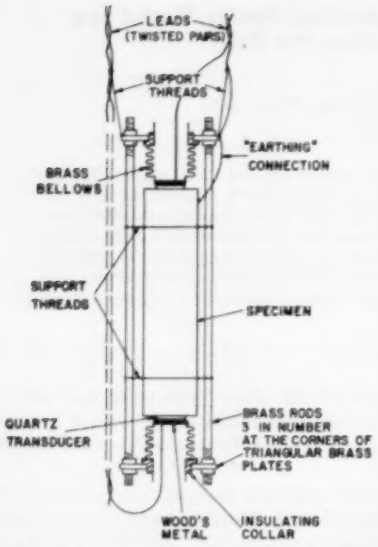


FIG. 2. Experimental arrangement for mounting the tin specimen.

linearity with the help of an input pulse fed from a source of constant amplitude through a calibrated attenuator; it was found to be as linear as the attenuator was good—adequately so for the purpose of the main part of this experiment. With the gain of the amplifier kept constant and with the input pulse to the specimen keeping constant amplitude, the height of the transmitted pulse—as seen on the oscilloscope—becomes a relative measure of the absorption of the sound in the rod; this height was measured on the photograph with the help of a traveling microscope.

III. EXPERIMENTAL RESULTS

The length of the specimen was originally 11.55 cm. It was found that the absorption in the temperature range 3.73–4.2°K was very heavy and the transmitted pulse was not very much greater than the noise level of the amplifier. This is shown in Fig. 3. However, below the superconducting transition (3.73°K) in the absence of a magnetic field (other than the earth's field) the amplitude of the transmitted signal increased considerably as the temperature was lowered. This is seen both in Fig. 3 and Fig. 4; in fact, by 1.52°K the pulse amplitude had increased to some 150 times the amplitude observed at 3.73°K.

A graph of pulse height (or amplitude) plotted against a function of $T/3.73$ (where T is the absolute temperature) approximating to $(T/3.73)^2$ produced a straight line (as shown by curve 1 of Fig. 5). The function of $T/3.73$ was that given by Lock, Pippard, and Shoenberg[10] for the temperature dependence of the critical magnetic field; this result may be coincidental as far as the function is concerned—it certainly was coincidental in that it appeared that the length of the specimen was accidentally chosen in such a way that a straight line resulted. It can readily be shown that if

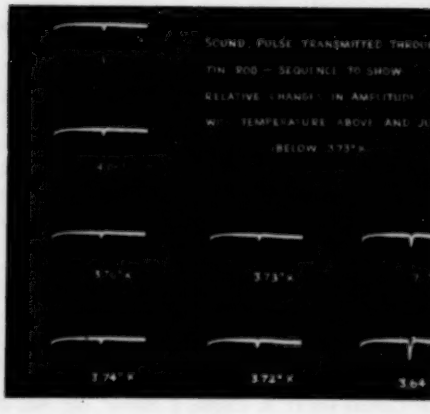

SOUND PULSE TRANSMITTED THROUGH TIN ROD — SEQUENCE TO SHOW RELATIVE CHANGE IN AMPLITUDE WITH TEMPERATURE ABOVE AND JUST BELOW 3.73°K
3.74°K
3.72°K
3.64°K

FIG. 3. Photographs of oscilloscope—3.64 to 4.15°K.

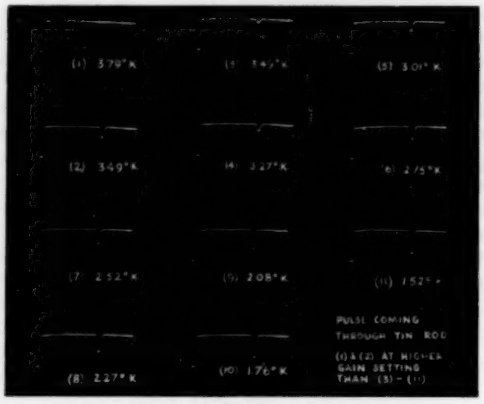


FIG. 4. Photographs of oscilloscope—1.52 to 3.79°K.

the normal law for sound absorption is obeyed, i.e., if

$$h_T = He^{-\alpha_T l}, \tag{1}$$

where h_T is the observed transmitted pulse height at temperature T, H the input pulse amplitude, α_T the amplitude absorption coefficient, in nepers/cm when l, the length of the specimen, is in cm, and if at some length l_1 the height h_{1T} is also given by

$$h_{1T} = h_{10}\phi(T)$$

where h_{10} is the transmitted pulse height at the absolute zero and $\phi(T)$ is some function of the absolute temperature, then at some other length l_2 the height h_{2T} will be given by

$$h_{2T} = h_{20}\psi(T),$$

[10] Lock, Pippard, and Shoenberg, Proc. Cambridge Phil. Soc. 47, 811 (1951).

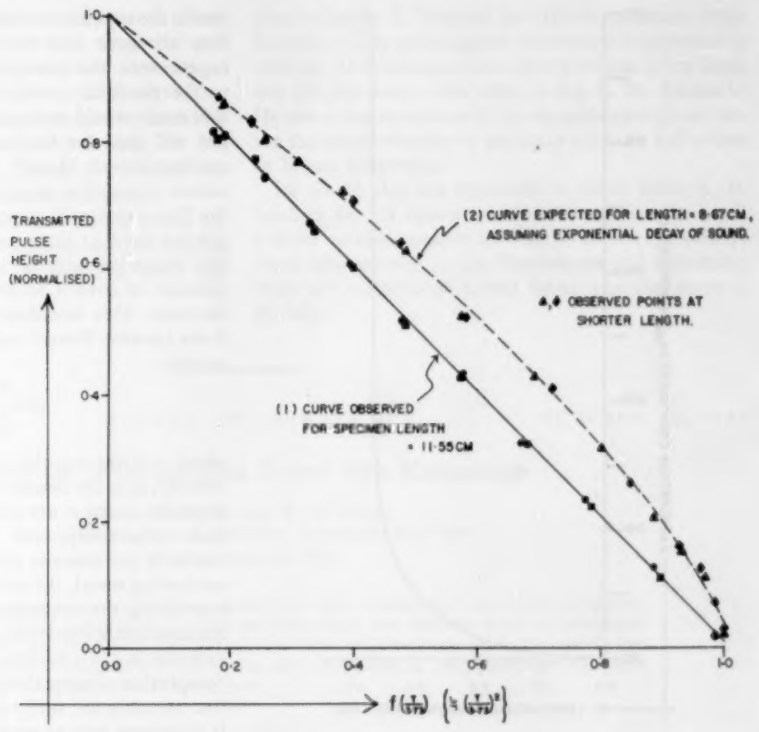


FIG. 5. Graphs of height of transmitted sound pulse (after normalization) plotted against $f(T/3.73)$ for both lengths of specimen.

where h_{20} is again the transmitted pulse height at the absolute zero and

$$\psi(T)=[\phi(T)]^{(l_2/l_1)}. \qquad (2)$$

(This result will hold whether or not there is absorption at the absolute zero of temperature.)

The length of the specimen was therefore reduced to 8.67 cm and, assuming the straight-line result for the longer length, the expected pulse height-temperature curve for the shorter specimen was calculated using Eq. (2) and has been incorporated in Fig. 5 as curve 2. Experimental points have been added to this figure and it can be seen that they lie reasonably close to this curve, thus confirming the negative exponential nature of the absorption and the earlier coincidence of the straight line.

It should be emphasized that all measurements were relative and that each set of experimental points required independent normalization before incorporation in the graph (Fig. 5).

The absolute value of the absorption at the absolute zero of temperature—or at any temperature—cannot be estimated from these results but, assuming that the extrapolation to the absolute zero is valid, then we can write that, for pure superconducting polycrystalline tin, the equation

$$\alpha_{T\,\min}=\frac{1}{11.55}\ln\left(\frac{1}{1-f(T/3.73)}\right), \qquad (3)$$

where

$$f\left(\frac{T}{3.73}\right)\doteq\left(\frac{T}{3.73}\right)^2,$$

and $\alpha_{T\,\min}$ is a minimum value for the absorption coefficient α_T [as defined by Eq. (1)], will be reasonably valid for all temperatures below but not very close to 3.73°K. Furthermore, $\alpha_{T\,\min}$ may differ from α_T by a constant undetermined quantity (equal to the value of α_T at 0°K) throughout this temperature range.

Equation (3) would predict an infinite absorption at the transition temperature. However, the observed value of $\alpha_{T\,\min}$ at this temperature was about 1.0 ± 0.4 nepers/cm and, using this, we can construct an $\alpha_{T\,\min}$-temperature graph as shown in Fig. 6.

The relatively high uncertainty in the value of $\alpha_{T\,\min}$ at the transition temperature arises from a discrepancy which appears when $\alpha_{T\,\min}$ is estimated from the different lengths of tin rod; the longer length gives a higher value than the shorter length. This may be due to slight nonlinearity of the amplifier which becomes important only when very small pulses are involved and is not important with larger pulses. The error involved here does not affect the general conclusions or the values for $\alpha_{T\,\min}$ below the transition—merely the point at which the absorption coefficient-temperature curve for tin above the transition temperature will break away from the very steep part of the curve of Fig. 6.

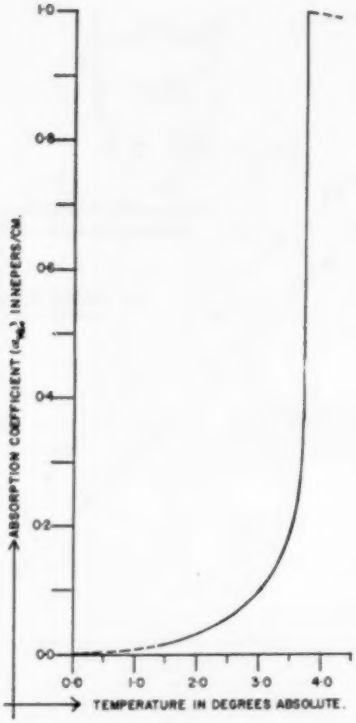


FIG. 6. Graph of the electronic absorption coefficient ($\alpha_{T\,min}$) plotted against absolute temperature.

IV. DISCUSSION AND CONCLUSIONS

It is clear that if superconductivity is a consequence of electron-lattice interaction, this change in absorption of sound must be in some way related, and reasonable qualitative explanations of the effect have already been published.[11,12] Theory so far[11–13] has only succeeded in accounting in a quantitative manner for the heavy absorption in the normal metal and has made no explicit predictions as to the variation with temperature of the absorption coefficient in the superconducting state—although there has been a prediction by Landau and Khalatnikov[14]; this however does not seem immediately applicable to these results as it discusses only the absorption to be expected in the superconducting state as a result of the disturbance by the sound wave of the equilibrium between superconducting and normal electrons; the experiments would seem to indicate that absorption by the normal electrons alone is perhaps the more important process. While the theories of absorption in the normal metal differ slightly in their approach, they all agree that for the frequency used in these experiments, the absorption coefficient is proportional to the electrical conductivity. (They show also that this result would not hold for an infinite conductivity but will hold for the usual experimentally observed conductivities.) Morse[13] and Kittel[12] both treat the normal absorption as a consequence of a relaxation of the Fermi surface in momentum space and Morse[13] has pointed out that the experimental results indicate that this relaxation must be very much speeded up by the presence of even a small number of superconducting electrons. This would certainly seem to be the case; if one assumes Morse's value for α in the normal metal, namely,

$$\alpha=\frac{4}{15}\frac{\omega^2 m E_0 \sigma}{\rho_0 c_0^3 e^2}$$

where ω is the angular frequency of the sound, c_0 its velocity, ρ_0 is the density of the metal, and σ is its conductivity, e and m are the charge and mass of the electron, respectively, and E_0 is the Fermi level, and further if one assumes this result to hold for the superconducting metal, the variation in the product $E_0\sigma$ (the σ involving the relaxation time) needed to account for the observed variation of α is too great to be correlated with existing two-fluid models of superconductivity on a "proportion of normal electrons" basis, (even allowing that one does not really know how E_0 and σ will vary). It is perhaps easy to see that collisions between normal and superconducting electrons might very well shorten the relaxation time (as they might involve momentum distributional changes for the normal electrons without energy changes), but no simple model has been found to which will fit the observations. The picture does however seem to suggest that the value of the electronic sound absorption coefficient (previously called $\alpha_{T\,min}$) in the superconducting state, being dependent on some different mechanism (possibly of an electron-electron interaction type, but not necessarily so and not necessarily involving a "two-fluid" concept), may be almost independent of its value in the normal state where the relaxation mechanism is very likely to be affected by those impurities (including lattice imperfections) which affect electrical conductivity values. However, further experimental work is needed to establish this plausible but uncertain point; it could not be expected to hold true when the value of normal electrical conductivity is low.

The discussion of these experiments has ignored absorption due to the lattice only, which has presumably to be added to the electronic absorption to obtain the total absorption. However, the variation with temperature of lattice absorption can reasonably be taken as small in the temperature range covered.

Experiments have been carried out on the effect of a magnetic field on the absorption; the results obtained will be the subject of a separate report, but it can be

[11] W. P. Mason, Phys. Rev. **97**, 557 (1955).
[12] C. Kittel, Phys. Rev. **98**, 1181(A) (1955).
[13] R. W. Morse, Phys. Rev. **97**, 1716 (1955).
[14] L. D. Landau and I. M. Khalatnikov, Doklady Akad. Nauk. S.S.S.R. **96**, 469 (1954). I am indebted to Dr. R. T. Beyer for drawing my attention to this paper and for preparing a translation; a few copies of this translation have been mimeographed and are available to those interested.

stated at this stage that existence of a hysteresis effect, as discovered by Bömmel,[2] has been confirmed and that in the normal metal, the absorption does depend strongly on magnetic field (so far applied only along the specimen). This effect was also originally found by Bömmel[15]; in the specimen described here, at the shorter length and at 3.9°K, about 200 oersted will double the amplitude of the transmitted pulse.

The author is very grateful for advice given by Dr. H. J. McSkimin of the Bell Telephone Laboratories in the making of the transducer-specimen bond and also to Dr. H. E. Bömmel for early information about his work. (This investigation was started independently without the knowledge that other work was being done, but the discovery of the effect is due to Dr. Bömmel.) He also acknowledges with thanks assistance given during the measurements by graduate students and others at Brown University.

He would also like especially to thank Dean R. B. Lindsay for his encouragement throughout this work and for making possible his visit to Brown University. He is also grateful to the Headmaster and Governing Body of Charterhouse School for an extended leave of absence.

[15] H. E. Bömmel (verbal communication).

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Properties of Germanium Doped with Manganese

H. H. Woodbury and W. W. Tyler
General Electric Research Laboratory, Schenectady, New York
(Received June 23, 1955)

The temperature dependence of the electrical resistivity and Hall coefficient in p- and n-type manganese-doped germanium crystals indicates that manganese introduces two acceptor levels in germanium at 0.16±0.01 ev from the valence band and 0.37±0.02 ev from the conduction band. The distribution coefficient for manganese in germanium is $(1.0\pm0.2)\times10^{-4}$. Comparison is made with other fourth-row metals (V, Fe, Co, and Ni) as impurities in germanium.

I. INTRODUCTION

SINGLE crystal germanium has been doped with manganese in a manner similar to the doping of germanium with Fe,[1] Co,[2] and Ni.[3] This paper will present the pertinent data related to the properties of manganese-doped germanium followed by a short comparison of the known properties of the fourth-row transition elements as impurities in germanium.

II. CRYSTAL PREPARATION

The details of the preparation of doped single-crystal germanium and the cryostat and assorted equipment used to take the measurements reported below are described in reference 1. In what follows the words crystal and sample will always refer to manganese-doped single crystal germanium unless noted.

Zone-melted manganese prepared by F. H. Horn of this laboratory, manganese obtained from Johnson-Matthey Company and manganese from Electro-Manganese Corporation, Knoxville, Tennessee, were all used with essentially the same results from each source.

In order to control the position of the Fermi level, the germanium melts were counterdoped by adding specific amounts of a single crystal Ge-As alloy. When used below, counterdoping will refer specifically to the addition of the Ge-As alloy either to compensate low-ionization energy acceptor impurities or to fill the levels introduced by the manganese.

[1] W. W. Tyler and H. H. Woodbury, Phys. Rev. **96**, 874 (1954); R. Newman and W. W. Tyler, Phys. Rev. **96**, 882 (1954).
[2] Tyler, Newman, and Woodbury, Phys. Rev. **97**, 669 (1955).
[3] Tyler, Newman, and Woodbury, Phys. Rev. **98**, 461 (1955).

III. EXPERIMENTAL RESULTS

A. Crystal Growing

It was found that when the pulling rate of the growing crystal was reduced from approximately 6 cm/hr to approximately 1.5 cm/hr, roughly 10 times as much Mn could be incorporated into the single crystal before the abrupt onset of lineage markings occurred.[1] These markings are apparently the result of small pockets of Mn or of a different Mn-Ge phase being incorporated in the growing crystal. With the growing rate equal to 1.5 cm/hr and a manganese concentration of 10^{19} to 10^{20} atoms/cc in the melt, the distribution coefficient (k_0) was $(1.0\pm0.2)\times10^{-4}$. The maximum solubility under the above conditions was approximately 10^{15} manganese atoms per cc. All crystals were grown in the (100) direction.

B. Electrical Properties

When manganese alone was used to dope the germanium melt, the crystals were always p-type. By counterdoping with a small amount of donor material, a low-lying level at 0.05 ev from the valence band and a second level at 0.16 ev from the valence band could be

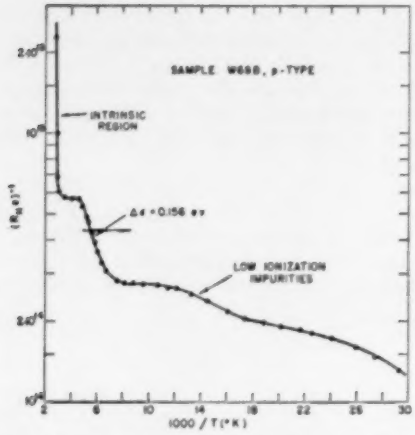


FIG. 1. P-type Mn-doped germanium crystal showing $(R_H e)^{-1}$ *vs* reciprocal temperature. In the extrinsic region, the ordinate is equal to the number of carriers (holes). The Mn introduces a level at 0.16 ev as shown by the change in the ordinate between 4 and 8 on the abscissa. The change in the number of carriers between 10 and 20 on the abscissa corresponds to an energy of 0.054 ev.

observed as shown in Fig. 1. The source of the 0.05-ev level could not be ascertained. It does not seem to be directly related to the manganese although it is not usually found in undoped crystals and was observed when each of the different sources of manganese were used. The level does not seem to be due to copper because of the higher activation energy measured than that reported in the literature[4] but this possibility has not been eliminated. Spectrographic analysis of the manganese doping material did not show any impurities which could be correlated with the 0.05-ev level. The analysis did indicate though that there could be sufficient Al impurity in the Mn to account for the lowest energy acceptor states (0.01 ev) found. These lowest states cannot be due to boron, a possibility that could not be eliminated by the spectrographic analysis. Boron has a distribution coefficient appreciably greater than one and hence it would quickly be taken up by the growing crystal, causing the first part of the crystal to be "more p-type" than the latter part. No such evidence of this type of variation along the length of the crystals was observed, indicating that there was less than 0.01 ppm of boron impurity in the manganese.

After these low-lying p-type "impurities" have been compensated, the crystals go to high resistance as they are cooled as shown in Fig. 2. These crystals are still p-type and the results definitely indicate that Mn introduces an acceptor level in the forbidden band 0.16 ev from the valence band.

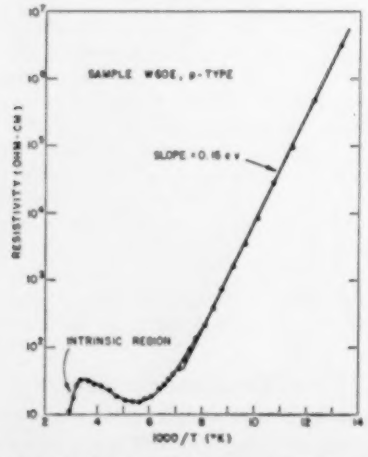


FIG. 2. Resistivity *vs* reciprocal temperature for a high-resistivity p-type Mn-doped germanium crystal.

[4] For example, see W. C. Dunlap, Jr., Phys. Rev. **96**, 40 (1954).

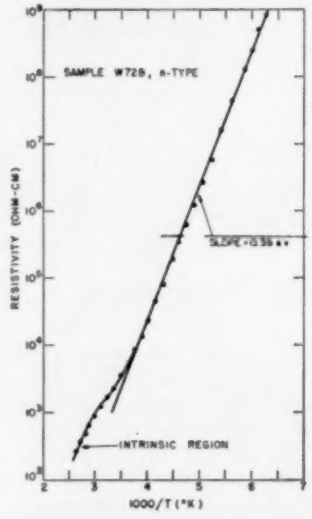


FIG. 3. Resistivity *vs* reciprocal temperature for a high-resistivity n-type Mn-doped germanium crystal.

Further counterdoping gives n-type material which goes to high resistance as shown in Fig. 3. The break in in the curve at $1000/T$ equal to approximately 3.5 marks the point between the extrinsic and the intrinsic regions. Still further counterdoping gives n-type material which does not go to high resistance when cooled as shown in Fig. 4. Because the characteristic energy of this n-type material is so near the middle of the forbidden band, it is impossible to unambiguously distinguish between the intrinsic and extrinsic regions in this case.

The levels at 0.16 ev and 0.37 ev have been directly related to manganese by varying the concentration of the manganese doping. Also, for a given concentration of manganese the amount of counterdoping necessary to fill the 0.16-ev level is the same, within the experi-

mental error of 20%, as that required to fill the 0.37-ev level. No such correlation with counterdoping is found for the 0.05-ev low-lying p-type level, and at the higher concentrations of manganese the number of 0.05-ev states is less than a third of the number of 0.16-ev states, as shown in Fig. 1. In investigating these levels, over 30 manganese-doped crystals have been grown, using varying amounts of counterdoping material.

The slope of log resistivity *versus* reciprocal temperature has been used as a measure of the ionization energy of the impurity states. In addition, for p-type samples that do not go to high resistance (such as shown in Fig. 1), direct computation of the ionization energy is possible since one knows from such curves the temperature at which the Fermi level is coincident with the impurity level. The average values obtained for these two levels are 0.16 ev$\pm$0.01 ev from the valence band for the one and 0.37 ev$\pm$0.02 ev from the conduction band for the other. The errors quoted are only a measure of the internal consistency of the data. Because of the manner of obtaining the energy from the temperature dependence of the resistivity, any linear dependence of the energy of the level with temperature would not be seen and the values quoted are taken to be the energy values of the levels at absolute zero.

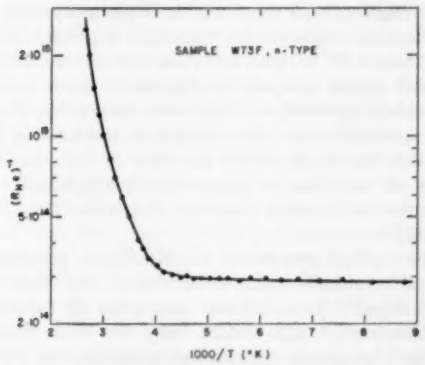


FIG. 4. N-type Mn-doped germanium crystal showing $(R_H e)^{-1}$ *vs* reciprocal temperature. In the extrinsic region, the ordinate is equal to the number of carriers (electrons).

The temperature dependence of the mobilities of these samples and others has been measured. In general, agreement is found between Mn-doped crystals and Fe-,[1] Co-,[2] and Ni-doped[3] crystals previously reported. However, because of the higher impurity concentration attained with Mn-doping, evidence of appreciable impurity scattering due to the manganese has been obtained and further work on this problem is being carried out.

C. Photoexperiments

Infrared photoconductivity studies show a spectral dependence qualitatively similar to those reported previously for gold-,[5] iron-,[1] cobalt-,[2] and nickel-doped[3] germanium crystals. Threshold energies for impurity photoconduction in both n- and p-type samples are consistent with the energy values obtained from resistivity studies. The n-type manganese-doped crystals show very large quenching effects and high photosensitivity.[6]

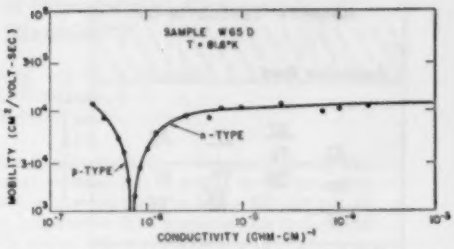


FIG. 5. Effective mobility *vs* conductivity of a p-type Mn-doped germanium crystal. The conductivity is changed by varying the amount of light falling on the sample.

Mobility measurements of optically excited carriers give results similar to the previously reported experiments on samples doped with Fe, Co, or Ni.[7] An example of such a measurement for a p-type Mn-doped sample is shown in Fig. 5. At a fixed temperature, the effective Hall mobility is plotted as a function of the conductivity which is changed by controlling the amount of light falling on the sample. The light is from a tungsten lamp at 3000°K and passes through a germanium filter held at the same temperature as the sample. The predominant effect of the light is to create electron-hole pairs but as shown in the figure, the holes are trapped since the sign of the Hall coefficient reverses when the conductivity is changed by less than a factor of three. In n-type samples no decrease in effective mobility is observed after an increase of conductivity by as much as a factor of 10^6. This again indicates that the holes created by the light are trapped, leaving only electrons to contribute to the conductivity.

The room temperature lifetime of p-type crystals has been measured using a spark light source and an oscilloscope to measure the decay of the photoconductivity. Within the experimental variations (a factor of 2) the room temperature lifetime is inversely proportional to the concentration of manganese impurity. For example, changing the number of manganese impurities from 0.3×10^{14} atoms/cc to 7×10^{14} atoms/cc causes a corresponding change in room temperature lifetime from 110 μsec to approximately 4 μsec. The product of the impurity concentration and the room temperature lifetime averages to 3×10^8 sec atoms/cc. No limiting lifetime values, i.e., lifetimes of heavily doped n- or p-type samples have been determined.

For n-type crystals, the behavior near room tempera-

[5] R. Newman, Phys. Rev. **94**, 278 (1954).
[6] R. Newman and H. H. Woodbury (to be published).
[7] W. W. Tyler and R. Newman, Phys. Rev. **98**, 961 (1955).

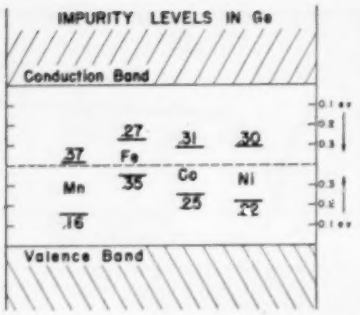


FIG. 6. Impurity levels introduced into germanium by various elements of the fourth row of the periodic chart. In addition to these elements, it is to be noted that vanadium is electrically inactive or is insoluble in germanium.

ture is quite different. At 60°C, the recovery times are roughly twice those for equivalent p-type material. As the samples are cooled to room temperature and below, the response time becomes longer and often nonexponential, showing evidence of hole trapping. Further work on the temperature dependence of photoconductivity decay is in progress.

IV. SUMMARY AND DISCUSSION

Figure 6 summarizes the observed energy values of the impurity levels introduced into germanium by transition elements of the fourth row of the periodic chart which have been investigated. The results of doping with iron, cobalt, and nickel are found in references 1, 2, and 3 respectively. Work on energy levels in germanium introduced by other impurity elements has been summarized by Burton.[8]

The evidence that the impurity elements noted in Fig. 6 introduce double acceptor levels in germanium is fairly conclusive. The evidence is direct from counterdoping experiments for all the elements except for iron. It is possible that for any one of the elements investigated, some acceptor impurity could have been present in just the right amount to make a possible donor level appear as an acceptor level. However, the striking similarity of the optical properties of germanium doped with these elements indicates that these levels are basically the same. And it would be extremely fortuitous if these elements all had impurities that would just compensate a donor-type level.

In addition to these elements, germanium has also been doped with vandium in a manner similar to that described above for manganese. Five vandium-doped crystals were grown with no observable effect on the resistivity or on the room-temperature lifetime. Lineage markings appeared on the crystal when the concentration of the vanadium in the germanium melt became approximately one-half of one percent with a pulling rate of approximately 2 cm/hr. Hall measurements from room temperature to 25°K were made on n-type and p-type vanadium-doped samples whose type was controlled by counterdoping. From analysis of these data, we can put an upper bound on the number of states introduced by the vanadium and thus on an effective distribution coefficient. This upper bound will depend on the energy of the assumed state. Assuming that each vanadium atom introduced either an acceptor or a donor hydrogen-like level, than the distribution coefficient for vanadium in germanium must be less than $3\cdot10^{-7}$ unless such levels were accidentally compensated by impurities in the vanadium. On the other hand, on the assumption that each vandium atom introduces a deep level, i.e., a level in the forbidden band greater than approximately 0.1 ev from either band, then the Hall measurements show that it is necessary that the distribution coefficient for vanadium in germanium be less than $2\cdot10^{-5}$. Hall mobilities were calculated and showed typical temperature dependent characteristics for undoped germanium. To account for the above facts, it is concluded that either vanadium is soluble in germanium but is electrically inactive, or that the solubility of vanadium in germanium is appreciably less than that of the other transition metals that have been studied.[9]

The optical properties of Mn-doped germanium crystals[6] are similar to those reported for Fe-,[1] Co-,[2] and Ni-[3] doped crystals. These impurities all introduce double-acceptor sites which may be made doubly charged by proper counterdoping during the crystal preparation. The presence of doubly-charged negative sites is the basis for hole trapping and the consequent high photosensitivity and quenching effects observed in n-type samples. The double acceptor model has also proven adequate for a qualitative understanding of the Hall coefficient studies of optically excited carriers,[7] studies of charged impurity scattering, and transient measurements.

[8] J. A. Burton, Physica **20**, 845 (1954).

[9] In addition to these observations, negative results are obtained when germanium is doped with the fourth column elements, tin and lead. On the basis of the reported value of 0.02 for the distribution coefficient of tin (see reference 8), crystals have been prepared containing more than 10^{19} tin atoms/cc. The presence of tin at this concentration has no observable affect on lifetime, resistivity, or mobility and it is concluded that tin is electrically inactive in germanium. These results were obtained using "C. P. Baker's Analyzed" tin. When "commercial high-purity" tin was used, n-type impurities were introduced into the germanium from impurities in the tin.

Electrical and Optical Properties of Intermetallic Compounds. IV. Magnesium Stannide*†

R. F. Blunt, H. P. R. Frederikse, and W. R. Hosler
National Bureau of Standards, Washington, D. C.
(Received July 8, 1955)

Measurements of electrical and optical properties of this semiconductor have been made over a wide temperature range. Results from conductivity and Hall effect measurements as well as optical absorption data indicate an energy gap of 0.33 ev at $T=0°K$. Mobility values were found to be $\mu_e=320$ cm²/volt-sec (for electrons) and $\mu_p=260$ cm²/volt-sec (for holes) at room temperature. Both parameters are proportional to $T^{-2.5}$. Appreciable photoconductivity is observed at 85°K and at 5°K.

INTRODUCTION

MOTT and Jones[1] point out in their discussion of Brillouin zones in crystals that for several compounds with the fluorite structure "the number of available valency electrons is just equal to the number of states within the zone" and therefore "we are led to expect a small conductivity." They list the following examples: Mg_2Sn, Mg_2Pb, Mg_2Si, Mg_2Ge, Li_2S, Na_2S, Cu_2S, and Be_2C. The experimental work that has been carried out on these compounds is, however, very limited. Two of the magnesium compounds (Mg_2Sn and Mg_2Pb) were investigated in 1948 by Robertson and Uhlig[2] who came to the conclusion that Mg_2Sn is a semiconductor with a forbidden energy gap of 0.26 ev, while Mg_2Pb showed metallic conduction.[3] Recently Busch and Winkler[4] measured the electrical conductivity and Hall coefficient of the compounds Mg_2Si, Mg_2Ge, and Mg_2Sn, and of mixed crystals of the latter two materials.[5] These investigators obtain for Mg_2Sn an energy gap of 0.36 ev at absolute zero. The difference between their result and that of Robertson and Uhlig is due to a different temperature dependence of the mobilities found by Busch and collatorators. The corresponding values of Mg_2Si and Mg_2Ge are listed as 0.77 ev and 0.74 ev, respectively.[4]

This paper deals in detail with the properties of Mg_2Sn.

EXPERIMENTAL PROCEDURES

Magnesium stannide (melting point 780°C[6]) was prepared by melting the two components together under a hydrogen atmosphere. The magnesium metal was obtained from two sources: Dominion Magnesium Limited (purity about 99.7%) and the Dow Chemical Company (purity better than 99.97%). Tin was supplied by the Vulcan Detinning Company (purity 99.997%). Thorough mixing in the molten state is an essential part of the preparation of this compound. The ingots which were pulled from the melt were very hard and brittle and showed a bluish color. These boules contained in most cases large single crystal areas and cleaved very easily along the (111)-plane exhibiting mirror-like surfaces. Contact with water covers the compound immediately with a dark layer quite analogous to the behavior of AlSb.[6]

The purity of the samples prepared in this way is remarkably high in spite of the heavy evaporation of magnesium during preparation. The compound has a tendency to grow crystals of nearly perfect stoichiometry and rejects excess magnesium as well as excess tin. Most crystals showed *n*-type conductivity; *p*-type ingots were obtained by doping with silver (*C*-22) or gold (*C*-23).

Samples for electrical measurements were ground to an approximate size of $1\times2\times7$ mm.

Several attempts to obtain good optical surfaces for transmission measurements by polishing with emery powder or alundum in alcohol were not very successful. The surfaces produced in this way showed large scattering. Excellent results were achieved by cleaving; all optical measurements reported in this paper were made on cleaved samples approximately $\frac{1}{2}$ mm thick.

X-ray analysis confirmed the fluorite structure of the crystals, the lattice constant being 6.7630 A[7] at 26°C.

Detailed descriptions of the equipment used for measuring Hall coefficient and resistivity,[8] optical absorption,[9] and photoconductivity[10] have been presented previously.

RESISTIVITY AND HALL EFFECT

The temperature dependence of the resistivity, ρ, and the Hall coefficient, R, of several *n*- and *p*-type

* Work supported by the Office of Naval Research.
† I, II, and III of this series have been published in Phys. Rev. **96**, 571–580 (1954).
[1] N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys* (Clarendon Press, Oxford, 1936).
[2] W. D. Robertson and H. H. Uhlig, Trans. Am. Inst. Mining Met. Engrs. **180**, 345 (1949).
[3] Their results do not exclude the possibility that Mg_2Pb might be a semiconductor with a very small energy gap.
[4] G. Busch and U. Winkler, Physica **20**, 1067 (1954); Helv. Phys. Acta (to be published).
[5] G. Busch and U. Winkler, Helv. Phys. Acta **26**, 579 (1954).
[6] Blunt, Frederikse, Becker, and Hosler, Phys. Rev. **96**, 578 (1954).
[7] Measurements made by the Microstructure Section of the National Bureau of Standards.
[8] R. G. Breckenridge and W. R. Hosler, Phys. Rev. **91**, 793 (1953).
[9] Breckenridge, Blunt, Hosler, Frederikse, Becker and Oshinsky, Phys. Rev. **96**, 571 (1954).
[10] H. P. R. Frederikse and R. F. Blunt, *Photoconductivity* (John Wiley and Sons, Inc., New York, to be published).

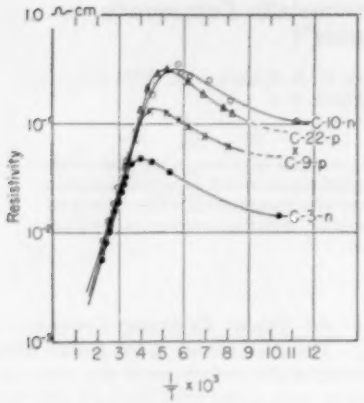


FIG. 1. Resistivity of various Mg_2Sn samples.

samples are given in Figs. 1 and 2. These samples contain between 2×10^{16} cm^{-3} and 3.5×10^{17} cm^{-3} impurities as calculated from Hall effect data.

Following the method indicated in reference 9 [Eq. (5)], one finds a ratio of electron to hole mobility of 1.23 (from results on samples *C*-9 and *C*-22). The variations of the mobilities with temperature are plotted in Fig. 3; both mobilities obey a $T^{-2.2}$ law.

In order to determine the value of the energy gap, two different procedures have been followed. For the *p*-type samples, it was possible to calculate the number of electrons (n) and holes (p) at every temperature from the resistivity and Hall coefficient, using the above-quoted value of the mobility ratio and the $T^{-2.2}$ temperature dependence. One can evaluate then

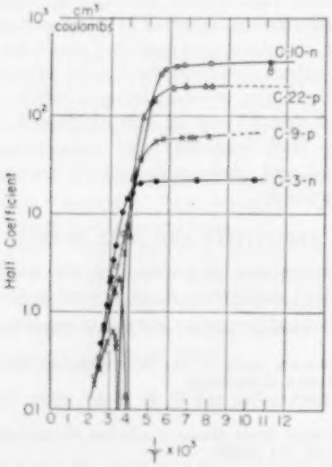


FIG. 2. Hall coefficient of various Mg_2Sn samples.

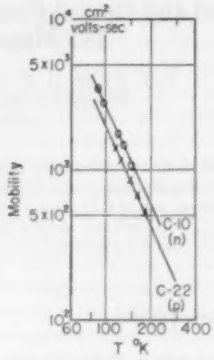


FIG. 3. Charge carrier mobilities of Mg_2Sn.

the product np which satisfies the following expression[11]:

$$np=4\left(\frac{2\pi m_e kT}{h^2}\right)^{\frac{3}{2}}\left(\frac{m_n^* m_p^*}{m_e^2}\right)^{\frac{1}{4}}\cdot e^{-E/kT}, \qquad (1)$$

where m_n^*, m_p^* = effective masses of electrons and holes respectively. Assuming that $E=E_0+\beta T$, it is clear that a plot of $\ln(np/T^3)$ *vs* $1/T$ should yield a straight line with a slope of E_0/k, where E_0 is the energy gap at absolute zero.

This procedure cannot be used for the *n*-type samples. In this case a graph showing $\ln(RT^{\frac{3}{2}})$ *vs* $1/T$ should produce a straight line, the slope being $E_0/2k$.

Plotting the experimental data in the manner indicated, one finds indeed a linear behavior in both cases; resulting values for the energy gap are given in the second column of Table I. Three of the four E_0 values agree fairly well; the average of these three is 0.33 ev.

The above-mentioned plot of $\ln(np/T^3)$ *vs* $1/T$ also enables us to determine the product of the effective masses, $(m_n^* m_p^*)/m_e^2$, from the intercept if the other remaining parameter $\beta(=dE/dT)$ is known. This quantity β may be evaluated from optical absorption

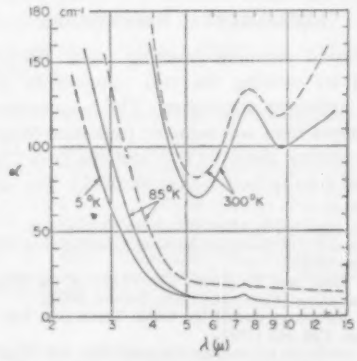


FIG. 4. Absorption spectra of *n*-type Mg_2Sn. - - - - *C*-10, ——— *C*-20.

[11] See reference 9 [Eq. (9a)].

data (see below); calculation then yields a value of 1.5 for the product $(m_n^* m_p^*)/m_e^2$. Combining this result with the relation[9]

$$\mu_n/\mu_p = (m_p^*/m_n^*)^{5/2} \quad (2)$$

leads to the following values of the effective masses: $m_n^* = 1.17 m_e$ and $m_p^* = 1.28 m_e$.

OPTICAL ABSORPTION

Infrared absorption spectra of four Mg_2Sn samples have been measured at three different temperatures. The results are given in Figs. 4 and 5.[12] Several features are immediately apparent. The absorption at 300°K is much larger than at low temperatures. This is due to the high number of intrinsic carriers (about 4×10^{17} cm^{-3} electrons and holes). Both n- and p-type samples show at room temperature a relatively strong absorption band with a maximum at about 7.5μ. No indication of this band was found in any of the spectral curves of the p-type samples at lower temperatures; a faint band is still observable in the spectra of the n-type samples at 85°K and at 5°K. The origin of this band (imperfections or lattice vibrations) is still an open question and can only be solved by means of additional experiments.

TABLE I. Values of the energy gap (ev) calculated from electrical and optical data.

	From electr. meas.	From optical absorption measurements			From photoconductivity measurements	
	$T=0°K$	$T=300$	$T=85°K$	$T=5°K$	$T=85°K$	$T=5°K$
C-3(n)	0.33[5]					
C-9(p)	0.28[5]					
C-10(n)	0.32[6]	0.23	0.29		0.28	0.30
C-20(n)		0.22	0.31	0.33	0.29	0.30
C-22(p)	0.34	0.21	0.30	(0.30)		0.30
C-23(p)		0.21	0.30	0.33		

A point of major interest is the value of the intrinsic energy gap as derived from transmission measurements. In the absence of any well-founded criterion, we have determined our values of the energy gap by extrapolating the absorption edge to α (absorption coefficient) = 0. The extrapolation consists of extending the linear part of the absorption edge towards the axis of abscissas; in this process considerable weight has been given to points with high α. Values for the energy gap obtained in this way are tabulated in Table I.[13] (Results on sample C-10 have to be regarded with much reservation due to the nonuniform thickness of the specimen.)

The shift of the absorption edge is -3.5×10^{-4} ev/deg between 85°K and 300°K.

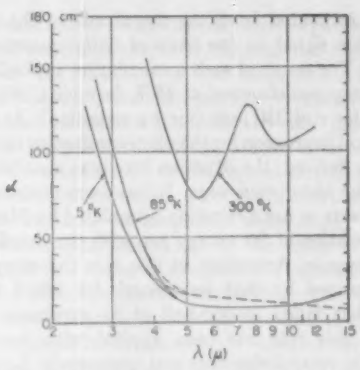


FIG. 5. Absorption spectra of p-type Mg_2Sn. —— C-22, - - - - C-23.

PHOTOCONDUCTIVITY

The photoconductive response of two Mg_2Sn samples is shown in Fig. 6. The ordinate, $\delta R/U$, represents the relative change of resistance ($\Delta r/r$) per incident photon (total flux or total number of photons = U). These measurements were made with a 0.5-mm slit opening of the monochromator at a modulation frequency of 510 cps. The signal at low temperature is several orders of magnitude above the noise; no photoconductivity could be detected at room temperature.

The photoresponse at 85°K is independent of the chopping frequency up to 510 cps, indicating a time constant smaller than 1 millisecond. The lifetime τ

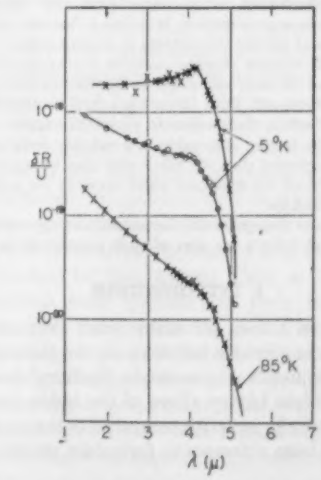


FIG. 6. Spectral response of two Mg_2Sn samples. × C−10, o C−22.

[12] The absorption coefficients are calculated on the basis of n (index of refraction) = 4.4; this value was obtained from preliminary reflectivity measurements.

[13] The very gradual change of slope of the absorption edge for low α makes an exact determination of the energy gap of Mg_2Sn very difficult. It has been suggested [G. G. Macfarlane (private communication)] that the rather long tail of the edge is due to indirect transitions between the valence and the conduction band [Hall, Bardeen, and Blatt, Phys. Rev. 95, 559 (1954)]. On the basis of this suggestion the energy gap would be considerably smaller. A critical evaluation of our electrical and optical data does not support such a small value of the energy gap.

may be estimated from the magnitude of the photoconductive signal on the basis of unity quantum efficiency.[14] The result of such a calculation applied to the largest response observed at 85°K (sample C-20) yields a value for τ of 250 μsec (for 4 μ radiation). As yet no theoretical expression for the photoconductive threshold has been derived; the situation here is as unsatisfactory as for the absorption edge. It has been customary in recent years to use a criterion formulated by Moss[15] for the correlation of the energy gap with the cutoff of the photoresponse. According to this rule the energy gap is determined by that wavelength for which the response has fallen to one-half of its maximum value. Indeed, this rule has been applied with reasonable success to several elements and compounds. Loferski,[16] however, has shown that this criterion leads to significant disagreement in the case of tellurium; extrapolating to absolute zero, one finds the photoconductive threshold of tellurium at 0.30 ev, while electrical measurements indicate a value for E_g of 0.34 ev. According to Loferski the energy gap is more closely connected with the maximum of the spectral distribution.

Identifying the width of the energy gap with either the maximum or the "break" of the photoconductive response, we find values of E as indicated in the last two columns of Table I. These figures are somewhat lower than the results from electrical and transmission data.

CONCLUSIONS

The compound Mg_2Sn appears to be a semiconductor, which obeys the "simple" theory rather well. Hall effect and conductivity data, as well as optical absorption data, indicate an energy gap of about 0.33 ev at absolute zero. The value deduced from the photoconductive threshold (4.2 μ) is, however, about 0.03 ev lower. An explanation of this discrepancy will probably have to await a better understanding of the correlation between optical transitions and band structure. The magnitude of the photoresponse is quite sizable at low temperatures.

The authors are greatly indebted to Mr. D. E. Roberts for preparing most of the crystals used.

[14] A. Rose, *Photoconductivity* (John Wiley and Sons, Inc., New York, to be published).
[15] T. S. Moss, *Photoconductivity in the Elements* (Butterworth Scientific Publications, London, 1952).
[16] J. J. Loferski, Phys. Rev. **93**, 707 (1954).

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Effect of Point Imperfections on the Electrical Properties of Copper. II. Thermoelectric Power*

F. J. Blatt
Physics Department, University of Illinois, Urbana, Illinois
(Received June 24, 1955)

The changes in the thermoelectric power of copper due to the presence of interstitials and vacancies have been calculated in the free electron approximation. It is found that the changes for concentrations of defects encountered in experiments on radiation damage due to massive charged particles are, at liquid nitrogen temperature, of the same magnitude as the absolute thermoelectric power of pure copper. Both interstitials and vacancies tend to reduce the absolute thermoelectric power of copper, and this effect should be readily observable in a suitably designed experiment. It is suggested that the effect may also be used to throw additional light on the processes which occur in the annealing of an irradiated sample.

The calculated changes in the thermoelectric property of copper due to Frenkel defects are also of such magnitude as to make adequate precautionary measures necessary whenever the thermocouple, which is used to measure the temperature of the specimen, is also in the beam of the massive charged particles. Unless the thermocouple is suitably screened from the beam, unreliable temperature measurement is likely to result. Curves showing the predicted temperature and concentration dependence of the thermoelectric power change due to Frenkel defects are presented.

The effect of small concentrations of arsenic in solid solution in copper on the thermoelectric power of copper has also been calculated by the same procedure. The calculated results are in satisfactory agreement with experiment, indicating that the results for interstitials and vacancies are probably of the correct magnitude.

I. INTRODUCTION

IT has been known for many years that impurities exert a considerable influence on the thermoelectric properties of metals. For example, Norbury[1] found that almost all dilute binary alloys of the noble metals are thermoelectrically negative compared to the pure metal. There have been attempts to formulate empirical rules relating the thermoelectric power of a dilute alloy to that of the pure metal; but these, apart from their failure to be satisfying from a fundamental point of view, are not universally applicable. For example, Kohler[2] is able to account for most of Norbury's results, but copper-manganese alloys form a notable exception to his additivity rules.

In the case of copper, at any rate, the change of

* Supported by the Office of Naval Research.
[1] A. L. Norbury, Phil. Mag. **2**, 1188 (1926).
[2] M. Kohler, Z. Physik **126**, 481 (1949).

thermopower with addition of impurities is most pronounced at very low temperatures, where changes in sign and in orders of magnitude have been observed.[3] For example, at 10°K the addition of 0.0026 atomic percent of Sn changes the absolute thermoelectric power from its value of about 0.3 μvolt/degree to −60 μvolt/degree. Other impurities give rise to similar astonishingly large effects. At these temperatures the anomalously large changes in thermoelectricity appear to be intimately related to the appearance of a resistance minimum, a phenomenon which has yet to be explained satisfactorily. In what follows, we shall not concern ourselves with this very low temperature region but restrict our discussion to that range of temperature in which the free electron approximation may, perhaps, be employed with more confidence, although even there a certain skepticism is justified in view of the many discrepancies between theoretically predicted and experimentally observed thermoelectric forces.

The thermoelectric power of alloys was analyzed recently by Domenicali and Otter,[4] in whose paper a good review of the experimental situation is also presented. Their analysis involves consideration of the change in the Fermi energy due to alloying as well as the details of scattering of conduction electrons by the impurity atoms. The energy dependence of the cross section for momentum transfer is expressed in terms of two parameters, α_1 and α_2, which are determined by fitting their expressions for the change in thermoelectric power to experimental observation. As they and Friedel[5] have pointed out, these parameters may also be derived theoretically from Mott's expression for the resistivity due to small impurity concentrations.[6] In this connection, however, it is well to bear in mind that Mott's formula is based on the use of the Born approximation; but, as shown by Fujiwara[7] and others, a more exact partial wave analysis leads to very different numerical results.

In the present discussion, we shall be concerned primarily with the effect of Frenkel defects on the thermoelectric power of copper. We shall also briefly consider the influence of arsenic in copper so as to be able to compare calculated with measured thermoelectric power changes and thereby obtain at least a rough estimate of the validity of our results for vacancies and interstitials.

The influence of Frenkel defects on the thermoelectric properties of metals is of interest for several reasons. In experiments on the effects of cyclotron or pile irradiation on solids the temperature of the sample is generally measured by means of a thermocouple attached to the specimen. The thermocouple, however, is calibrated prior to irradiation; since a portion of the couple is generally also in the flux of the high-energy particles, the reliability of the temperature measurement is suspect if the thermoelectric properties of the couple are substantially altered by the presence of vacancies and interstitials. Of course, it is only that arm of the couple, such as the copper arm of a copper-constantan thermocouple, which is a pure metal that is likely to have its properties significantly altered; since the residual resistance of the alloy is already quite high, it is rather unlikely that a concentration of about 0.2 atomic percent of Frenkel defects, encountered in the strongest cyclotron irradiation used at this time, would influence its thermoelectric behavior very much.

Changes in thermoelectricity due to the presence of vacancies and interstitials might also present themselves as yet another tool for investigating the annealing spectrum of an irradiated specimen. On the basis of previous results,[8] it is probable that both types of imperfections contribute almost equally to the resistivity of copper, and it is, therefore, practically impossible to distinguish annealing of interstitials from that of vacancies on the grounds of observed resistivity changes alone. On the other hand, as the following results indicate, interstitials probably exert a significantly larger influence on the thermoelectric power than do vacancies, so that simultaneous resistance and thermoelectric power measurements on an irradiated copper sample could lead to results which may throw additional light on the annealing of radiation damage in copper.

Finally, thermoelectricity is more sensitive than resistivity to the details of the scattering process of conduction electrons in metals. For the latter, all that matters is the total cross section for momentum transfer, while the former is related to the dependence of this cross section on the electron energy. Calculated thermoelectric power changes are, then, also more sensitive, generally, to the model which one chooses to represent the impurity, and comparison with experiment may permit the elimination of inferior models.

In calculating $\Delta S/S$, the relative change of the absolute thermoelectric power of copper due to the presence of imperfections, we shall assume that the Fermi level of the conduction electrons of copper remains unaffected by concentrations of imperfections of the order of magnitude encountered in radiation damage, that is 0.2 atomic percent or less. This assumption is justified for two reasons. First, at these small concentrations each imperfection finds itself in an environment of the pure metal, and the overlap of screening charge around the imperfections is negligible. Under these conditions, as shown by Friedel,[7] the Fermi level remains unchanged. Furthermore, even for slightly larger concentrations, for which a small change in the Fermi level might be expected, the correction which

[3] D. K. C. MacDonald, Physica **19**, 841 (1953).
[4] C. A. Domenicali and F. A. Otter, Phys. Rev. **95**, 1134 (1954).
[5] J. Friedel, J. phys. radium **14**, 561 (1953).
[6] N. F. Mott, Proc. Cambridge Phil. Soc. **32**, 281 (1936).
[7] H. Fujiwara, J. Phys. Soc. Japan **10**, 339 (1955). J. Friedel, Advances in Phys. **3**, 446 (1954).
[8] F. J. Blatt, Phys. Rev. **99**, 1708 (1955). Hereafter referred to as I.

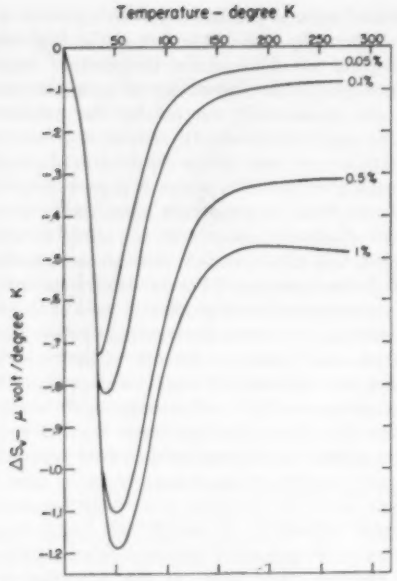


FIG. 1. ΔS_v, the changes of the absolute thermoelectric power of copper resulting from concentrations of 1.0, 0.5, 0.1, and 0.05 atomic percent of vacancies, as functions of temperature.

this introduces into the expression for $\Delta S/S$ is quite small.

In the free electron approximation the expression for the absolute thermoelectric power is

$$S=(\pi^2k_0^2T/3e)(d\log\rho(\epsilon)/d\epsilon)_{\epsilon_0}, \tag{1}$$

where e is the electronic charge, k_0 Boltzmann's constant, T the absolute temperature, $\rho(\epsilon)$ the resistivity for electrons of energy ϵ, and the logarithmic derivative is to be evaluated at the Fermi energy ϵ_0.

When an imperfection is introduced, raising the resistivity of the metal to $\rho+\Delta\rho$, the accompanying change in the thermoelectric power is

$$\Delta S=\frac{\pi^2k_0^2T}{3e}\left[\frac{d\log(\rho+\Delta\rho)}{d\epsilon}-\frac{d\log\rho}{d\epsilon}\right]_{\epsilon_0} \tag{2}$$

and the fractional change, $\Delta S/S$, is given by

$$\Delta S/S=\frac{-(1-\Delta x/x)}{(1+\rho/\Delta\rho)}, \tag{3}$$

where

$$x=\epsilon_0\left(\frac{d\log\rho}{d\epsilon}\right)_{\epsilon_0},\quad \Delta x=\epsilon_0\left(\frac{d\log\Delta\rho}{d\epsilon}\right)_{\epsilon_0}. \tag{4}$$

The quantity x is simply related to the absolute thermoelectric power of the pure metal and is determined from experimental observation. The change in resistivity due to the addition of imperfections is also known from observation in many instances, or, in the case of vacancies and interstitials, has been calculated recently by the present author and others.[8,9] The only unknown quantity in Eq. (3) is Δx, and it has the virtue of being, for a given metal, a function only of the type of imperfection present and does not depend on the concentration or temperature. Thus, once this last quantity has been evaluated, the concentration and temperature dependence of $\Delta S/S$ can be ascertained without further calculation.

II. RESULTS

We make the same assumptions as in I, namely, (1) The free electron approximation is valid. (2) The scattering potential associated with the imperfection has spherical symmetry. (3) Matthiessen's rule is valid. (4) The effect of relaxation of the lattice about the imperfection is negligible.

We make use of the additional assumption, discussed earlier, that the Fermi level of the free electron gas remains unchanged.

The resistivities $\Delta\rho_i$ and $\Delta\rho_v$ due to interstitials and vacancies, respectively, have been calculated in I and are

$$\Delta\rho_i=1.4\ \mu\text{ohm-cm/atomic percent},$$

$$\Delta\rho_v=1.3\ \mu\text{ohm-cm/atomic percent}.$$

There remains the problem of determining Δx in each case. This was done by evaluating the resistivities $\Delta\rho_i$ and $\Delta\rho_v$ at energies $\epsilon_0+\delta\epsilon$ and $\epsilon_0-\delta\epsilon$, where $\delta\epsilon$ was

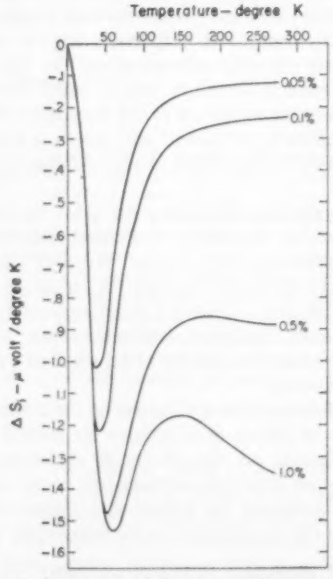


FIG. 2. ΔS_i, the changes of the absolute thermoelectric power of copper resulting from concentrations of 1.0, 0.5, 0.1, and 0.05 atomic percent of interstitials, as functions of temperature.

[9] P. Jongenburger, Appl. Sci. Research **B3**, 237 (1953); Nature **175**, 545 (1955).

selected arbitrarily to be $\epsilon_0/20$, and approximating the derivative in Eq. (4) by the ratio of the differences, i.e.

$$\Delta x \simeq 10\{\log[\Delta\rho(\epsilon_0+\delta\epsilon)]-\log[\Delta\rho(\epsilon_0-\delta\epsilon)]\}. \quad (5)$$

The scattering potentials from which the resistivities were calculated were derived from the Hartree self-consistent potentials in the manner discussed in I.

The quantity x was obtained from the measured absolute thermoelectric power of copper,[10] and for the resistivity the values given by Meissner[11] were used. The results are shown in Figs. 1–4.

In order to gain some confidence in the numerical results, the effect of arsenic on the thermoelectric power of copper was also determined in a similar manner. Arsenic was chosen because of the impurities considered in I it is the only one for which the effect on thermoelectricity is known from experiment.[12]

Taking the same scattering potential as in I, Δx was found to be 2.1, which compares favorably with the value 2.0 deduced from observation. Abelès[13] has suggested the use of a square-well potential adjusted so as to satisfy the Friedel sum condition [see Eq. (5) of I], and this procedure was also carried out with the result $\Delta x = 1.2$. Both procedures are thus seen to be satisfactory, leading to results of the correct magnitude, although, perhaps fortuitously, the former appears to be slightly better.

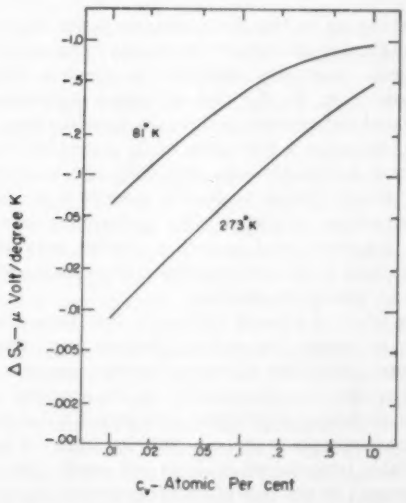


FIG. 3. ΔS_v, the change of the absolute thermoelectric power of copper resulting from vacancies, as a function of vacancy concentration, at 81°K and 273°K.

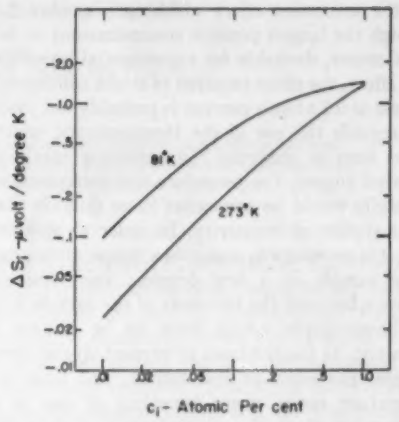


FIG. 4. ΔS_i, the change of the absolute thermoelectric power of copper resulting from interstitials, as a function of interstitial concentration, at 81°K and 273°K.

III. DISCUSSION

Figures 1 and 2 show a rather strong temperature dependence of the change in thermoelectric power due to Frenkel defects. This temperature variation arises first of all from the temperature variation of the denominator of Eq. (3) through the approximately linear dependence of ρ on T. In addition, there is the dependence of S on T which, ideally, should be linear so that the numerator of Eq. (3) should be independent of temperature. However, below about 150° K the thermoelectric power of copper deviates significantly from the linear temperature variation predicted by Eq. (1), and this departure from the predictions of the free electron theory leads to the pronounced minimum of ΔS.

Vacancies and interstitials both tend to decrease the thermoelectric power of copper, a circumstance which would make it more difficult to distinguish between these two imperfections on the basis of thermoelectric measurements. Nevertheless, such separation may still be possible by careful experimental techniques since interstitials should be considerably more effective than vacancies in reducing the thermoelectric power of the specimen. The largest change in thermoelectricity should occur in the vicinity of liquid nitrogen temperature, although even at higher temperatures one should be able to detect small changes in S with sensitive equipment. Of course, the major limitation on experimental observation at high temperatures is the small number of defects which are likely to be encountered. Annealing studies on irradiated copper have shown that most of the damage has already annealed out before room temperature is reached.[14]

Figures 3 and 4, which show the concentration dependence of ΔS, clearly exhibit, at the lower tempera-

[10] J. Nyström, Arkiv Mat. Astron. Fysik **A34**, No. 27 (1948).
[11] W. Meissner, *Handbuch der Experimentalphysik* (Akademische Verlagsgesellschaft, Leipzig, 1935), Vol. 11, Part 2, p. 51.
[12] Borelius, Keesom, Johansson, and Linde, Leiden Comm. **206b** (1930).
[13] F. Abelès, Compt. rend. **237**, 796 (1953).
[14] Cooper, Koehler, and Marx, Phys. Rev. **97**, 599 (1955).

ture, the saturation effect which sets in when $\Delta\rho \gtrsim \rho$. Although the largest possible concentrations of defects are, of course, desirable for experimental investigation of the effect, the effort required to attain concentrations in excess of 0.1 atomic percent is probably not justified.

As regards the use of the thermoelectric effect described here in studying the annealing spectrum of irradiated copper, the procedure and interpretation of the results would be somewhat more difficult than in similar studies of resistivity. In order to observe the effect, it is necessary to maintain a temperature gradient in the sample of a few degrees; this temperature difference between the two ends of the sample forming the thermocouple would have to be known quite accurately. As the specimen is warmed up the temperature gradient must be maintained; and thus, in the temperature range where annealing of one or more defects takes place, the concentration of these defects in the sample will not be uniform, annealing occurring more rapidly at the higher temperature end. Furthermore, even if no annealing were to take place over a certain temperature range, ΔS will nevertheless exhibit more or less drastic variations, as predicted by the curves of Figs. 1 and 2. Thus, deducing the annealing spectrum from an anlysis of ΔS *versus* T for an irradiated sample would be considerably more involved than interpreting the changes in resistivity over the annealing range.

We have so far refrained from speculating about the accuracy of the results. A certain skepticism is justified by the fact that the free electron model predicts the wrong sign for the absolute thermoelectric power of pure copper. Apart from this, the nonlinear temperature dependence of S in the range below about 150° K throws further doubt on deductions which are based on the use of Eq. (1). Nevertheless, the agreement found for the calculated and observed effect of arsenic on the thermoelectric power of copper would indicate that our results are probably of correct magnitude.

Finally, we have not considered an effect which has recently been brought to focus, namely the "phonon contribution" to the thermoelectric power. This has been investigated by Herring for semiconductors[15] and by Ter Haar and Neaves for metals.[16] The conclusion, in the case of metals, is that in the temperature region around $\theta/2$ the drag on the charge carriers due to the phonon current may be substantial and could, perhaps, account for the increase of the thermoelectric power of copper with decreasing temperature which is observed between 150 and 80° K. It is reasonable that the presence of imperfections in the lattice would reduce the phonon contribution to S because the scattering of conduction electrons associated with these imperfections reduces the importance of the electron-lattice interaction in determining the relaxation time, and also because the imperfections will scatter phonons and thus tend to restore an equilibrium phonon distribution. It is, therefore, quite possible that the actual decrease of the thermoelectric power due to Frenkel defects is enhanced by the phonon contribution in that temperature range where this effect is important.

To date there is no experimental evidence which either confirms or disproves the conclusions reached here. The only investigations on the effects of Frenkel defects on the thermoelectric properties of metals known to the writer are those of Andrew, Jeppson, and Yockey[17] and of Jamison and Blewitt.[18] Both groups of workers reported no observable changes. However, Andrew *et al.* performed their experiment at high temperatures. It is, therefore, not surprising that the density of defects was inadequate to give rise to measurable changes. Jamison and Blewitt subjected their samples, which were cooled to liquid nitrogen temperature, to a total neutron flux of 5×10^{16}. Neutrons are, regrettably, much less effective than charged massive particles in producing lattice defects; the concentration of defects achieved by Jamison and Blewitt was probably about 2×10^{-4} atomic percent, and again the negative result which they reported might have been anticipated. Evidently, further experimental investigations are needed in order to compare the present results with observation.

IV. CONCLUSIONS

The change in the thermoelectric power of copper due to Frenkel defects and due to small concentrations of arsenic have been calculated in the free electron approximation. In the case of arsenic agreement of calculated and experimental results is satisfactory. The predicted change in thermoelectricity caused by Frenkel defects is sufficiently large, especially at low temperatures, that it should be readily observable for defect concentrations attainable with cyclotron irradiation. It is suggested that annealing studies may throw further light on the processes which occur during warm-up of an irradiated specimen.

The effect of Frenkel defects on the thermoelectric power of copper also makes adequate precautionary measures imperative whenever thermocouples are used to determine the temperature of the specimen in a radiation damage experiment. It is probable that unless the thermocouple is properly shielded from the beam, unreliable temperature readings will result. The error will depend on the type of couple employed and on the fraction of it exposed to the radiation. Because of their intrinsic high residual resistance, a couple formed of suitable alloys is less likely to cause trouble than one in which one or both arms consist of pure metals.

[15] C. Herring, Phys. Rev. **96**, 1163 (1954).
[16] D. Ter Haar and A. Neaves, Proc. Roy. Soc. (London) **A228**, 568 (1955).
[17] Andrew, Jeppson, and Yockey, Phys. Rev. **86**, 643 (1952).
[18] R. E. Jamison and T. H. Blewitt, Rev. Sci. Instr. **24**, 474 (1953).

X-Ray Lattice Constant of Barium Oxide*

ROBERT J. ZOLLWEG†
Physics Department, Cornell University, Ithaca, New York
(Received June 21, 1955)

Precision measurements of the x-ray lattice constant of pure BaO were made at temperatures between 8°K and 435°K with a back reflection symmetrical focussing camera. Room temperature measurements of colored BaO indicate a contraction of the lattice associated with the presence of excess barium.

I. EXPERIMENTAL PROCEDURES

THE low-temperature symmetrical focusing camera used for these measurements consisted of a liquid helium cryostat similar to that described by Duerig and Mador[1] with the lower end modified so as to contain the x-ray camera. The BaO powder specimen was held between two beryllium foil sheets clamped in a copper block and bolted to the lower end of the liquid helium container of the cryostat. An aperture in the liquid nitrogen shield permitted the x-ray beam to strike the specimen and the diffracted x-rays to return to the photographic film at room temperature.

The BaO samples were either polycrystalline specimens grown by the vapor deposition technique[2] or part of a compressed cylinder of BaO that had been held at about 1500°C for some hours in a vacuum as part of the crystal growing process (the presence of facets indicating some recrystallization takes place). These were ground in a mortar and passed through a sieve in a dry box desiccated with porous BaO and liquid nitrogen. Impurities were probably about 0.05% calcium, 0.02% strontium, 0.002% magnesium, and 0.001% molybdenum. The BaO powder was sprinkled onto one sheet of Be foil containing a couple of drops of ceresin wax[3] dissolved in benzene (dried with porous BaO), the second sheet of Be foil was placed on top and the "sandwich" clamped in the copper holder. The ceresin wax furnished adequate protection for the BaO from the atmospheric water vapor during the short periods required to mount the specimen in the cryostat and for changing the photographic films, even in the absence of the top sheet of beryllium foil. The cryostat was evacuated by a forepump during the exposures. A 0.003-in. copper-constantan thermocouple junction was imbedded with the BaO at the center of the "sandwich." Most specimens were not annealed after grinding because of the extreme instability of BaO in the presence of water vapor. However, one specimen was annealed by heating to 800°C in a quartz tube evacuated by a forepump and a liquid nitrogen trap. No change in the lattice constant was detected as a result of annealing.

Cobalt K_α radiation ($K_{\alpha 1}=1.78892$ A, $K_{\alpha 2}=1.79278$ A) was used and diffraction lines corresponding to $h^2+k^2+l^2$ equal to 36, 35, and 32 were obtained. Exposures were about 30 minutes at 25 kv and 15 ma. The films were calculated making use of Cohen's method[4] and gave an average standard deviation of about 0.0005 A. With the low-temperature camera the greatest source of error was due to the powder specimen being off the camera circumference and Cohen's correction appropriate for this error was employed. No correction for refractive index was made in view of the current disagreement as to whether one should be employed for powder diffraction patterns.

II. RESULTS FOR STOCHIOMETRIC BaO

Measurements on six films taken with the low-temperature camera gave a value of the lattice parameter of stochiometric BaO at 25°C of 5.53912±0.00013 A. This result was checked by six films taken with a commercial camera which yielded a result 5.53915 ±0.00013 A.

Measurements at other temperatures were made using liquid nitrogen, liquid helium, or dry ice and acetone as coolants or a heated oil bath in the cryostat liquid helium container. The results of measurements of four specimens at different temperatures are shown in Fig. 1. The liquid nitrogen and liquid helium measurements with one polycrystalline sample (open triangles) are evidently in error for some unknown reason (the three other liquid nitrogen measurements are in good agreement). The curve in the figure is that obtained by fitting the data with the equation[5]

$$\frac{a-a_{0^\circ K}}{a}=\frac{\gamma}{3vK}U_T,$$

where γ is Grüneisen's number, K is the bulk modulus, v is the molar volume, and U_T is the thermal lattice energy less the zero-point energy. $a_{0^\circ K}$ was estimated to be 5.5240 A and the constant of proportionality determined by using the lattice parameter value for 25°C. The curve shown in the figure results if the lattice energy is expressed as the sum of a Debye

* This work was supported in part by the Office of Naval Research.
† Now at the Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania.
[1] W. H. Duerig and I. L. Mador, Rev. Sci. Instr. **23**, 421 (1952).
[2] Sproull, Dash, Tyler, and Moore, Rev. Sci. Instr. **22**, 410 (1951).
[3] A. S. Eisenstein, J. Appl. Phys. **17**, 434 (1946).
[4] M. U. Cohen, Rev. Sci. Instr. **6**, 68 (1935); Z. Krist. **91**, 288, 306 (1936). E. R. Jette and F. Foote, J. Chem. Phys. **3**, 605 (1935).
[5] C. Zwikker, *Physical Properties of Solid Materials* (Interscience Publishers, Inc., New York, 1954).

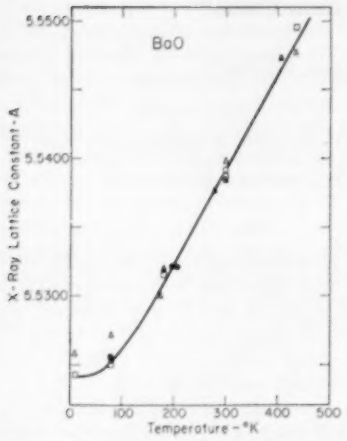


FIG. 1. X-ray lattice constant for several specimens of stochiometric BaO at various temperatures. Triangles represent two polycrystalline specimens grown by the vapor deposition technique, squares represent annealed sample, black dots are specimen from the heated compressed cylinder of BaO and open circle is the accurate room temperature value.

function with characteristic temperature of 173°K plus an Einstein function with characteristic temperature of 332°K (the function corresponding to that which Anderson[6] states best fits his BaO specific heat data), or if the lattice energy is expressed as a single Debye function with characteristic temperature of 280°K. The results are not sufficiently precise to distinguish between these functions though they lend support to Anderson's specific heat data by being consistent with it.

The linear expansion coefficient for BaO between 200°K and 400°K is found to be $12.8\pm0.9\times10^{-6}$/°K. This value is somewhat smaller than the value $17.8\pm2.1\times10^{-6}$/°K reported by Eisenstein[3] between room temperature and 875°C.

III. COLORED BaO RESULTS

BaO crystals are colored by annealing in an atmosphere of barium (blue BaO).[7] X-ray lattice constant measurements were made at room temperature on two relatively lightly colored crystals prepared at this laboratory by Timmer and on two densely colored crystals used by Sproull, Bever, and Libowitz[8] (two films each were made with three of these specimens). The samples were prepared in the same way as the stochiometric BaO and were not annealed after grinding. Figure 2 shows the resultant x-ray lattice parameter *vs* the optically determined density of absorption centers, assuming an oscillator strength of $\frac{1}{2}$ and using Smakula's formula (chemical measurements yielded somewhat greater concentrations of excess barium). There is a considerable contraction of the lattice in the case of the densely colored crystals. If one were to assume a linear contraction of the lattice with percentage of excess barium, one would obtain the relation $\Delta a/a=-0.1$ (atomic percent excess Ba).

These results are consistent with the view that the blue coloration is associated with oxygen defects in the lattice in the form of F' centers.[8] A contraction of the lattice around the F' center might be expected because of the reduction in the repulsive interaction when the oxygen ion is replaced by the F' center. A contraction is observed, for instance, when a chloride ion replaces a bromide ion in KBr.[9] In addition, Burstein, Oberly, and Davisson[10] have concluded from measurements of the pressure dependence of color center absorption in alkali halides that there is a displacement of ions near an F center toward the vacant site.

BaO crystals grown from solution in molten barium are also colored (red BaO).[11] Chemical analysis by Libowitz indicated an excess barium by weight of about 0.1%. Kane[12] made optical absorption measurements of these crystals and found the density of absorption centers to be about 8×10^{18} cm^{-3} assuming an oscillator strength of $\frac{1}{2}$ and using Smakula's formula. Measurements of three specimens of the red BaO yielded values of the lattice parameter of 5.5351 ±0.0007 A, 5.5337±0.0010 A, and 5.5328±0.0009 A at 25°C.

The lattice contraction observed with red BaO was smaller than that observed for an equivalent amount of excess barium in the blue crystals. The appearance of some contraction, however, implies that the red coloration is not caused by interstitial barium since interstitials would be expected to cause a considerable

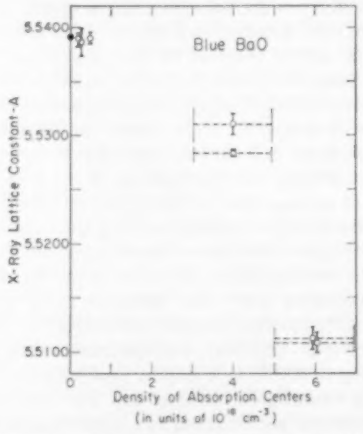


FIG. 2. The x-ray lattice constant of blue BaO *vs* the optically determined density of absorption centers.

[6] C. T. Anderson, J. Am. Chem. Soc. **57**, 429 (1935).
[7] W. C. Dash, Phys. Rev. **92**, 68 (1953).
[8] Sproull, Bever, and Libowitz, Phys. Rev. **92**, 77 (1953).
[9] C. R. Berry, J. Appl. Phys. **24**, 658 (1953); F. Oberlies, Ann. Physik **87**, 238 (1928).
[10] Burstein, Oberly, and Davisson, Phys. Rev. **85**, 729(A) (1952).
[11] G. G. Libowitz, J. Am. Chem. Soc. **75**, 1501 (1953).
[12] E. O. Kane, J. Appl. Phys. **22**, 1214 (1951).

lattice expansion. The smaller lattice contraction may be caused by aggregation, perhaps in the form of colloids. The fact that there is any contraction at all points to one of two possibilities:

(1) F' centers in equilibrium with other defects; or

(2) less lattice strain associated with the centers responsible for the red absorption, as might be the case for defects intermediate in size between single oxygen vacancies and colloids.

In this connection, it should be emphasized that red BaO crystals are grown at about 900°C, while blue crystals are colored at about 1100°C.

ACKNOWLEDGMENTS

The author is indebted to Professor R. L. Sproull for his continued guidance in this work and to Professor Emeritus C. C. Murdock for the use of some of his x-ray equipment.

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Classical Theory of Cyclotron Resonance for Holes in Ge

J. M. Luttinger and R. R. Goodman
University of Michigan, Ann Arbor, Michigan
(Received June 24, 1955)

Cyclotron resonance line shapes for holes have been calculated in the classical limit. The magnetic field was taken in the (001) direction. Because the cyclotron resonance frequency depends on the component of k parallel to the magnetic field (k_H), the line is shifted and broadened as compared to the simple theory with $k_H=0$. The shift of the maximum is about 3%, and the broadening (which is asymmetric) is about 40% for Ge, if $\omega\tau$ is taken as 7.5.

THE shape of cyclotron resonance absorption lines for Ge has been calculated treating the holes as classical particles. Quantum effects[1] become important at low temperatures and will be considered in a separate publication. It is easily shown that for a constant magnetic field H, the component of the quasi-momentum in the direction of H (k_H) is a constant of the motion. Thus for H in the (001) direction (the only case considered here), k_z is a constant. The equations of motion for k_x and k_y are then given by

$$\begin{aligned} \dot{k}_x &= \partial\mathcal{H}/\partial k_y \\ \dot{k}_y &= -\partial\mathcal{H}/\partial k_x, \end{aligned} \tag{1}$$

where for holes

$$\mathcal{H}=\frac{eH}{c}(\bar{A}k^2\pm[\bar{B}^2k^2+\bar{C}^2(k_x^2k_y^2+k_y^2k_z^2+k_z^2k_x^2)]^{\frac{1}{2}}). \tag{2}$$

The plus and minus refer to the light and heavy holes, respectively. The constants $\bar{A}$, $\bar{B}$, $\bar{C}$ are those reported by Dresselhaus, Kip, and Kittel[2] as A, B, C. One sees from Eq. (1) that k_x and k_y are canonically conjugate variables. The equations of motion for k_x and k_y will depend parametrically on k_z, and there will of course be a thermal distribution of k_z.

In order to estimate the effect of the thermal distribution of k_z a transport theory with a constant relaxation time τ was used. The theory is essentially that of Van Vleck and Weisskopf.[3] The formalism of the theory is similar to that used by Karplus and Schwinger.[4]

The calculations were performed starting with the following expression for the absorbed power per unit volume ρ:

$$\mathcal{P}=Ne\{\langle v_x(t)\rangle E_x(t)\}_{Av} \tag{3}$$

where E_x refers to an electric rf field in the [100] direction. N is the number of holes per unit volume and $v_x(t)$ is the velocity of a hole in the x direction. The brackets $\langle\ \rangle$ indicate an averaging over the time of collision and an averaging over the Boltzmann distribution. The braces refer to an average over time. Since k_x and k_y are canonically conjugate a new set of canonical coordinates P and Q, where $\mathcal{H}=P$, can be found. This leads to a considerable simplification. In terms of these new variables the averaging procedure gives

$$\mathcal{P}=\tfrac{1}{2}Ne^2E^2\tau\int_{-\infty}^{+\infty}dk_z\int_{P_0}^{\infty}dP\rho\beta\sum_{n=-\infty}^{+\infty}\frac{|v_x(n,P)|^2}{1+\tau^2(\omega+\omega_0 n)^2}. \tag{4}$$

ρ is the Boltzmann factor

$$\rho=\frac{\exp\left(-\dfrac{\beta c}{eH}P\right)}{\Delta},$$

where

$$\Delta=\int_{-\infty}^{\infty}dk_z\int_{P_0}^{\infty}dP\int_0^{2\pi/\omega_0}dQ\exp\left(-\frac{\beta c}{eH}P\right),$$

[1] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).
[2] Dresselhaus, Kip, and Kittel, Phys. Rev. **98**, 368 (1955).
[3] J. H. Van Vleck and V. F. Weisskopf, Revs. Modern Phys. **17**, 227 (1945).
[4] R. Karplus and J. Schwinger, Phys. Rev. **73**, 1020 (1948).

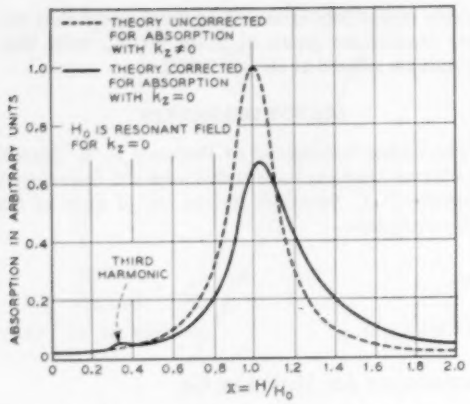

THEORY UNCORRECTED FOR ABSORPTION WITH $k_z \neq 0$
THEORY CORRECTED FOR ABSORPTION WITH $k_z = 0$
H_0 IS RESONANT FIELD FOR $k_z = 0$
THIRD HARMONIC
ABSORPTION IN ARBITRARY UNITS
$x = H/H_0$

FIG. 1. Cyclotron resonance line shape for holes in Ge.

and $\beta=1/kT$. P_0 is determined by the minimum value of $\mathcal{H}$ for a given k_z. ω is the frequency of the rf field and $\omega_0=\omega_0(k_z,P)$ is the frequency of the periodic motion of k_x and k_y for a given P and k_z. The $v_x(n,P)$ are the Fourier coefficients of a Fourier expansion of $v_x(P,Q)$ in terms of $e^{in\omega Q}$. It is easily shown that $v_x=-(c/eH)\dot{k}_y$. Thus, if k_y is expressed as a Fourier series

$$k_y=\sum_{n=-\infty}^{+\infty} C_n e^{in\omega_0 t}$$

and if η is defined as $\eta=P/k_z^2$, Eq. (3) becomes (after an integration over k_z)

$$\mathcal{P}=\frac{3eE^2\tau c}{4H\delta(\eta_0)}\int_{\eta_0}^{\infty} d\eta\eta^{-5/2}\omega_0(\eta)\sum_{n=-\infty}^{+\infty}\frac{n^2|C_n|^2}{1+\tau^2(\omega+n\omega_0)^2}, \quad (5)$$

where

$$\delta(\eta_0)=\int_{\eta_0}^{\infty} d\eta\eta^{-\frac{1}{2}}\omega_0^{-1}(\eta),$$

and η_0 is the smallest value of $\mathcal{H}$ for $k_z=1$.

The C_n and $\omega_0(\eta)$ can in principle be determined exactly by integration of the equations of motion for k_x and k_y. In fact, the results may be expressed as elliptic integrals of the third kind. These exact solutions turn out to be quite impractical for computation, and we have only used them for checking purposes. The results reported here were obtained by means of an analog computer, which integrated the equations of motion of k_x and k_y directly. The values of $\bar{A}$, $\bar{B}$, and $\bar{C}$ taken were those reported by Dresselhaus *et al.*[2] The final integration over η was performed numerically. A variational method was also used to determine the values of C_n and ω_0. There is good agreement between the results of the variational method and the results given by the computer.

In order to determine the values of $\bar{A}$, $\bar{B}$, and $\bar{C}$, Dresselhaus, Kip, and Kittel[2] also used classical methods; however, they used the approximation that $k_H=0$.[5] The results of calculations using $k_H=0$, and of our theory may be seen in Fig. 1 (where we have taken $\omega=1.5\times10$ in. rad/sec and $\omega\tau=7.5$). The corrected curve may be seen to have its maximum shifted by about 3% from that of the $k_H=0$ curve.[6] In addition an asymmetrical broadening of about 40% occurs. It should be mentioned that Zeiger[7] has reported similar calculations. He has, however, treated the asphericity of the energy surface as a perturbation on the main spherical term, and in addition limited himself to the neighborhood of resonance. His results seem to be in substantial agreement with those reported here.

With the above method it is possible to redetermine the constants $\bar{A}$, $\bar{B}$, and $\bar{C}$ more accurately. The change in their values will be not more than a few percent, and, at present, many other experimental and theoretical uncertainties overwhelm this correction.

It is also possible to calculate the effect of extra resonances due to higher harmonics. From Fig. 1 it can be seen that the third harmonic contribution is quite small for H in the [001] direction. The fifth harmonic was found to be much smaller. Further calculations are being made for H in the [111] direction, and for Si.[8]

ACKNOWLEDGMENTS

The authors wish to thank Professor J. Sellars for his very considerable help with the computations done on the analog computer.

[5] These authors mention an attempt to correct for the $k_H=0$ assumption [see discussion following their Eqs. (79) and (80)]. No indication of their method is given, and, especially for Si, their corrections seem somewhat high to us.

[6] The larger the $\omega\tau$, the less the maximum is shifted.

[7] H. J. Zeiger, Phys. Rev. **98**, 1560(A) (1955). Also private communication.

[8] In Si the effects discussed here will certainly be larger because of the greater asphericity of the energy surface.

Considerations on Double Exchange

P. W. ANDERSON, *Bell Telephone Laboratories,* Murray Hill, New Jersey*

AND

H. HASEGAWA, *University of Tokyo, Tokyo, Japan*

(Received June 21, 1955)

Zener has suggested a type of interaction between the spins of magnetic ions which he named "double exchange." This occurs indirectly by means of spin coupling to mobile electrons which travel from one ion to the next. We have calculated this interaction for a pair of ions with general spin S and with general transfer integral, b, and internal exchange integral J.

One result is that while the states of large total spin have both the highest and lowest energies, their average energy is the same as for the states of low total spin. This should be applicable in the high-temperature expansion of the susceptibility, and if it is, indicates that the high-temperature Curie-Weiss constant θ should be zero, and $1/\chi$ *vs* T a curved line. This is surprising in view of the fact that the manganites, in which double exchange has been presumed to be the interaction mechanism, obey a fairly good Curie-Weiss law.

The results can be approximated rather well by a simple semiclassical model in which the spins of the ion cores are treated classically. This model is capable of rather easy extension to the problem of the whole crystal, but the resulting mathematical problem is not easily solved except in special circumstances, e.g., periodic disturbances (spin waves).

I. INTRODUCTION

ZENER[1] has proposed a mechanism of electron spin interaction which he calls "double exchange." He suggested that this mechanism is responsible for ferromagnetism in the mixed-valency manganites of perovskite structure, such as $(La_xCa_{1-x})(Mn_x^{III}Mn_{1-x}^{IV})O_3$.[2] Zener points out that the extra electron on the Mn^{III} can travel back and forth between the two Mn ions, and that its spin will couple with those of both ion cores. This obviously leads to a spin coupling of some kind, which he shows is probably ferromagnetic. As Zener points out, if the transfer does not change the electron's spin, and if exchange between it and the ion is so large that it cannot be on an ion of the wrong spin, then the electron can move if the ions are parallel, and cannot move if they are not parallel. One expects therefore a parallel coupling of the order of magnitude of the transfer integral which causes the electron's motion.

We have investigated this mechanism in greater detail and with considerably more general assumptions than reference 1. We were interested in seeing whether it differed in any observable aspects from ordinary exchange. The result is indeed qualitatively different, as we could have anticipated from the fact that there is an extra degree of freedom, namely the motion of the electron, to be considered.

The results can be understood rather easily on a simple semiclassical model, which we shall give in the main body of the paper. The rigorous calculation will be given in an Appendix. Further, this simple model can be extended far enough so that we can see some of the characteristics of the problem which must be solved for an entire solid, although mathematical solution of this problem is practically impossible.

One of the major differences from ordinary exchange is that there are two levels for every spin arrangement of the ion cores, one as high in energy as the other is low. If exchange is small one of these corresponds to a spin-up traveling electron, the other to spin-down; while if exchange is large one is the symmetric, one the antisymmetric orbital state. Thus the *average* energy is not a function of total spin, and the series calculation of susceptibility[3] fails to give a second term, so that θ vanishes in the high-temperature Curie-Weiss law. This effect is not observed.[4]

II. THE MODEL; CONSIDERATION OF THE TRANSFER PROCESS

The model we use is shown in Fig. 1. We denote the magnetic atoms by Mn and some neutral intermediary atom by O (thinking of course of the case of the manganites). It is assumed that the O atom has a closed shell, the two magnetic atoms a number of d-electrons appropriate to give them each a spin S as well as one extra electron which may, of course, equally well be

FIG. 1. Model for double exchange.

* Part of this work was done while this author was a Fulbright Lecturer at the University of Tokyo. I should like to express my gratitude for the hospitality and support of the University and the Fulbright Commission.

[1] C. Zener, Phys. Rev. **82**, 403 (1951).

[2] G. H. Jonker and J. H. Van Santen, Physica **16**, 337, 599 (1950).

[3] J. H. Van Vleck, J. Chem. Phys. **5**, 320 (1937).

[4] G. H. Jonker and J. H. Van Santen (private communication).

TABLE I. Labeling of wave functions in double exchange.

	Mn_1	O	Mn_2
A. Orbitals			
Mobile d-electron	d_1		d_2
Closed O shell		p	
Fixed d-electrons	d_1' etc.		d_2' etc.
B. Configurations			
I	$d_1' d_1$	p^2	d_2'
II	d_1'	p^2	$d_2' d_2$
III	$d_1' d_1$	p	$d_2' d_2$

on one or the other. The wave functions and configurations considered are labeled in Table I. We assume from the first that there is a unique orbital d_1 or d_2 into which the extra electron wishes to go; our results are valid under more general conditions[5] but this will give the qualitative picture of the process which we wish to convey. We then have a problem of configuration interaction between the two degenerate configurations I and II with the electron on the one atom or the other. Zener assumed that the intra-atomic exchange integral coupling the electron's spin to S, i.e., $J_{dd'}$, was infinitely large, so that he excluded states with $S_{\text{tot}}=S-\frac{1}{2}$ from his configurations; we shall not make this assumption at first, although we believe this to be the most probable case. However, for the sake of simplicity we assume that the Mn atom core always has spin S; thus its internal exchange integrals are assumed infinitely large.

In order for there to be any coupling between the two Mn atoms, the extra electron must be able to get from one to the other. The transfer process must occur through the intermediary of the O atom for our picture. (Of course, there is nothing in our calculations requiring this, but most cases to which it is reasonable to apply these concepts do have a magnetically neutral atom present.) The transfer is most likely to occur in one of two ways. The first is that suggested by Zener: via the exchange type integral

$$J^*=\int d\tau\psi_{d1}(1)\psi_p(1)H\psi_p(2)\psi_{d2}(2), \tag{1}$$

which is however negligibly small (since the Mn atoms do not overlap) unless the p and d wave functions used are not orthogonal, in which case it is equal to

$$\begin{aligned} J^*&=S_{p-d}\int\psi_pV\psi_d d\tau \\ &=S_{p-d}H_{p-d}, \end{aligned} \tag{2}$$

where S is the overlap integral, V the averaged potential.[6] This will then be of roughly the same magnitude as the ordinary $p-d$ exchange integral for nonorthogonal wave functions. J^* has simply the effect of moving the electron from atom 1 to atom 2, i.e., configuration I to II, without change of spin, as can be seen by Serber's method[6] (see Appendix I). If orthogonal wave functions are used we must as usual go to higher order configuration interactions to find a mechanism, but when we do can rely somewhat more on our results. The configuration which must be introduced is that in which a single p electron has been transferred from the O to the "other" Mn, i.e., configuration III. This is connected to the initial configuration I by the ordinary transfer integral H_{p-d} of (2); then this same transfer integral can connect it with II, since I and II are symmetrical relative to III. The details of this process are worked out also in Appendix I, where we prove that this too is a simple transfer, the electron ending on the second ion with the same spin direction from which it started on the first. Both processes give transfer integrals of the rough order of magnitude 0.1 v, while J, the intra-atomic exchange integral, is of order 1 v.

III. SEMICLASSICAL CALCULATION OF DOUBLE EXCHANGE

The preceding section showed that the primary transfer mechanisms which may occur in double exchange are both of a type which tend simply to transfer a single electron without change of spin from atom 1 to atom 2. We now proceed to calculate the energies of the states which follow from configurations I and II when we take into account this transfer integral (which for simplicity we shall call simply b) and also the fact that the electron's spin is coupled to the spin of the ion on which it finds itself by an intra-atomic exchange integral J. In the case of ions with less than half-filled shells, J tends to make the spins parallel; for more than half-filled shells, and if the odd electron falls into the d shell and not into an s function, we have an "effective" exchange integral J which is infinite and causes antiparallelism. However, for the present problem the sign of J itself is irrelevant.

Our procedure from here on will be to treat the problem semiclassically, that is, as if the two ions had very large spins S so that one could assign them definite directions in space and a definite angle relative to each other. Two nearly trivial modifications of this procedure lead to the exact answer as derived in Appendix II.

When the electron is on the first atom, it has two states, of energies

$$E_1=\pm JS,$$

lined up parallel and antiparallel to the spin $\mathbf{S}_1$. On atom 2 it also has these two states, but lined up in the direction of $\mathbf{S}_2$. We know, however, that the electron, when transferred, goes only into the parallel state. There are two possible ways of correcting for the difference in direction: either to transform the energies

[5] These are either (a) the orbits are sufficiently quenched that S, $S+\frac{1}{2}$ and $S-\frac{1}{2}$ are unique states, and any other orbital states are of high energy, in which case our treatment in Appendix II is obviously valid; or else (b) J is large compared to the transfer integral and states S and $S+\frac{1}{2}$ are adequately quenched. We then ignore all $S-\frac{1}{2}$ states.

[6] R. Serber, Phys. Rev. **45**, 461 (1934).

on atom (2) into a coordinate system parallel to $\mathbf{S}_1$, or to transform the transfer matrix elements so that they refer correctly to the different directions. The latter procedure is somewhat simpler.

Let us label the two electronic spin functions referred to the direction of $\mathbf{S}_1$ by α and β, and those referred to $\mathbf{S}_2$ by α', β'. The eigenstates on atom 1 are $d_1\alpha$ and $d_1\beta$. These have the energies (if we make the first quantum-mechanical modification of using correct exchange energies)

$$\begin{aligned} E(d_1\alpha) &= -JS, \\ E(d_1\beta) &= J(S+1), \end{aligned} \tag{3}$$

for J ferromagnetic in sign. The eigenstates on (2) are

$$\begin{aligned} d_2\alpha' &: E = -JS, \\ d_2\beta' &: E = +J(S+1), \end{aligned} \tag{4}$$

while the transfer matrix elements are

$$\begin{aligned} &(d_1\alpha|H|d_2\alpha) = b, \\ &(d_1\alpha|H|d_2\beta) = 0, \text{ etc.} \end{aligned} \tag{5}$$

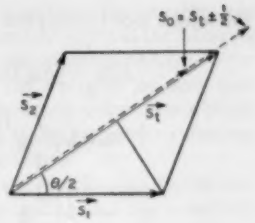


FIG. 2. Addition of spins in double exchange.

What we must do is to express α and β in terms of α' and β'. This transformation is well known,

$$\begin{aligned} \alpha &= \cos(\theta/2)\alpha' + \sin(\theta/2)\beta', \\ \beta &= -\sin(\theta/2)\alpha' + \cos(\theta/2)\beta', \end{aligned} \tag{6}$$

where θ is the angle between $\mathbf{S}_1$ and $\mathbf{S}_2$.

We can now immediately write down the secular equation:

$$|H-E| = 0$$

$$H = \begin{array}{c} \\ d_1\alpha \\ d_1\beta \\ d_2\alpha' \\ d_2\beta' \end{array} \begin{array}{c} \begin{array}{cccc} d_1\alpha & d_1\beta & d_2\alpha' & d_2\beta' \end{array} \\ \begin{vmatrix} -JS & 0 & b\cos\theta/2 & b\sin\theta/2 \\ 0 & J(S+1) & -b\sin\theta/2 & b\cos\theta/2 \\ b\cos\theta/2 & -b\sin\theta/2 & -JS & 0 \\ b\sin\theta/2 & b\cos\theta/2 & 0 & J(S+1) \end{vmatrix} \end{array}. \tag{7}$$

The secular equation reduces to

$$\begin{aligned} &\{(\tfrac{1}{2}J-E)^2 - [J(S+\tfrac{1}{2}) + b\cos\theta/2]^2 - b\sin^2\theta/2\} \\ &\times\{(\tfrac{1}{2}J-E)^2 - [J(S+\tfrac{1}{2}) - b\cos\theta/2]^2 - b^2\sin^2\theta/2\} = 0, \end{aligned}$$

which has the solutions

$$\begin{aligned} E &= \tfrac{1}{2}J \pm \{[J(S+\tfrac{1}{2}) \pm b\cos\theta/2]^2 + b^2\sin^2\theta/2\}^{\frac{1}{2}} \\ &= \tfrac{1}{2}J \pm [J^2(S+\tfrac{1}{2})^2 + b^2 \pm 2Jb(S+\tfrac{1}{2})\cos\theta/2]^{\frac{1}{2}}. \end{aligned} \tag{8}$$

This is the complete solution except that we must evaluate $\cos\theta/2$. In the semiclassical case this can be evaluated by noting that (see Fig. 2)

$$\frac{|\mathbf{S}_1+\mathbf{S}_2|}{2S} = \frac{S_t}{2S} = \cos\theta/2 \cong \frac{S_0}{2S}, \tag{9}$$

where this equation defines S_t. Now in the semiclassical case we cannot meaningfully distinguish between S_t and S_0, the total spin including the odd electron (equal to $S_t \pm \frac{1}{2}$), so this is the complete answer in terms of this theory, as noted in the last equality of (9).

In Appendix II, we shall show that the only modification introduced by exact calculation is to replace (9) by the result, equivalent to terms of order $1/S$,

$$(\cos\theta/2)_{\text{exact}} = (S_0+\tfrac{1}{2})/(2S+1). \tag{10}$$

The result for the intermediate case $J \cong b$ is not of very great interest. It will suffice to discuss the two limiting cases, (A) $J \gg b$, and (B) $J \ll b$.

A. $J \gg b$.—This is the case considered by Zener. Here we can ignore the eigenvalues having a + sign on the square root as having very high energies (i.e., ignore $d_1\beta$ and $d_2\beta'$). We then neglect $b^2\sin^2\theta/2$ in the square root and

$$\begin{aligned} E &\cong -JS \pm b\cos\theta/2 \\ &= -JS \pm b[(S_0+\tfrac{1}{2})/(2S+1)]. \end{aligned} \tag{11}$$

In case J has antiferromagnetic sign the result is unchanged except for a constant. S_0 can take on every value from $\frac{1}{2}$ to $2S+\frac{1}{2}$, so that $\cos\theta/2$ runs from $1/(2S+1)$ to 1. For each of these values there is one high-energy solution corresponding to each low-energy solution. Physically, what happens is that the effective transfer integral between the only two possible states for the electron is reduced by the ratio $(S_0+\frac{1}{2})/(2S+1)$, and the eigenstates are essentially the symmetrical and antisymmetrical combinations of the localized orbitals with the electron's spin parallel to $\mathbf{S}_1$ or $\mathbf{S}_2$, respectively.

B. $J \ll b$.—Here we again must neglect + values of the square root, and, adding a negligible term inside

the square root to complete the square, we get

$$E\cong-\frac{J}{2}-[b\pm J(S+\tfrac{1}{2})]\cos\theta/2$$
$$=-b+\frac{J}{2}\pm\frac{J(S+\frac{1}{2})(S_0+\frac{1}{2})}{2S+1} \qquad (12)$$
$$=-b+\frac{J}{2}\pm\frac{J}{2}(S_0+\tfrac{1}{2}).$$

Again, for every value of S_0 there are two states, one high and one low. These correspond to spin up or down for the traveling electron, relative to the total spin $\mathbf{S}_1+\mathbf{S}_2$. Again the sign of J is immaterial.

IV. DISCUSSION OF RESULTS

The results (8), (11), and (12) of the last section are quite different from the usual form of exchange interaction. They are of interest not only because they refer to a mechanism which may be occurring in certain substances but also because this is one of the simplest problems of spin interaction in which orbital degeneracy enters in an essential way. It is, in fact, only because of the extra degree of freedom allowed by the electron's motion that we can obtain anything other than an $\mathbf{S}_1\cdot\mathbf{S}_2$ interaction in this simple case.

Two differences are apparent: first, the linear variation with total spin S_0 in both (11) and (12), whereas $\mathbf{S}_1\cdot\mathbf{S}_2$ varies quadratically; and second, the fact that for every S_0 there are two states, not one, which in both (11) and (12) are symmetrically disposed about the average energy.

The first difference does not seem to be very essential physically. However, the second has at least a very probable significance. The susceptibility of a paramagnetic substance can always be written[7]

$$\chi=\langle\mu^2\rangle_{\rm Av}/3kT, \qquad (13)$$

where μ is the magnetic moment, which for a system with orbital quenching is

$$\boldsymbol{\mu}=g\beta\mathbf{S}, \qquad (14)$$

g being the g-factor and β the Bohr magneton. Thus χ is proportional to $\langle S^2\rangle_{\rm Av}/kT$. We may then take the average with a Boltzmann distribution

$$\langle S^2\rangle_{\rm Av}=\frac{\sum_{S_0,n}S_0(S_0+1)e^{-E_n(S_0)/kT}}{\sum e^{-E_n(S_0)/kT}}, \qquad (15)$$

which at high temperatures may be expanded into

$$\langle S^2\rangle_{\rm Av}=\frac{\sum_{S_0}S_0(S_0+1)\sum_n\left(1-\frac{E_n(S_0)}{kT}+\frac{E_n^2(S_0)}{2k^2T^2}\cdots\right)}{\sum_{S_0,n}\left(1-\frac{E_n(S_0)}{kT}+\frac{E_n^2(S_0)}{2k^2T^2}\cdots\right)}.$$

[7] J. H. Van Vleck, *Theory of Electric and Magnetic Susceptibilities* (Oxford University Press, London, 1932), p. 40.

If we wish only the first two terms we note that $\sum_n E_n(S_0)$ is independent of S_0, since the average energy for all S_0 is the same. Then up to terms in $1/T^2$

$$\langle S^2(T)\rangle_{\rm Av}=\langle S^2(\infty)\rangle_{\rm Av},$$

and we may write

$$\chi=C/T+(\text{terms in } 1/T^2). \qquad (16)$$

The corresponding expansion for ordinary exchange gives

$$\chi_{\rm ordinary}=C/T+C\theta/T^2+\cdots\cong C/(T-\theta), \qquad (17)$$

where θ is a temperature of the order of magnitude of the Curie temperature, i.e., such that $k\theta$ is comparable to the total interaction. Thus we suspect that a double-exchange ferromagnet will have a $1/\chi-T$ plot as shown in Fig. 3, as compared to a ferromagnet, and also to a typical ferrimagnet.

Of course, all these conclusions are based on the simplest case of only two ions with one electron travelling between them. In the real solid there are a very large number both of ions and of electrons, and there are $(2S+1)^N$ states for the ions and N for an electron for every one of these. We can for the sake of simplicity assume that the number of electrons is relatively small and ignore their interactions, but even the calculation of the states of a single electron plus the ion system is very complex.

With the confidence gained from Sec. III in the semiclassical treatment of the ions, however, we can at least draw some qualitative conclusions in our two cases A and B.

In case A, $J\gg b$, the transfer integral between any

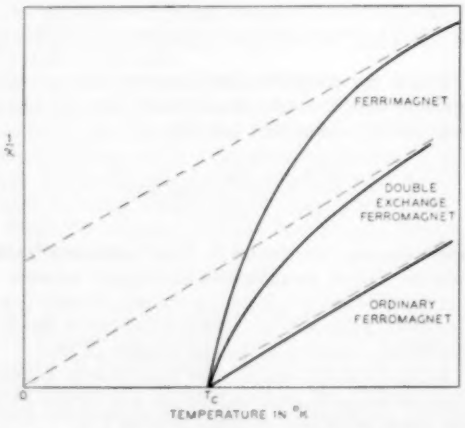


FIG. 3. Qualitative susceptibility curves for ferrimagnet, ferromagnet, and double exchange ferromagnet.

two ions depends on $\cos\theta/2$ for the angle θ between their two spins. For a random, paramagnetic type of spin arrangement we have then a problem of electron motion in a random alloy, which has been the subject of much attention but has not been solved in the way needed here. This case of variable transfer integral is also not the usual one. However, it is still to be expected that for each arrangement there is created a band of electronic states, as many above as below the average arrangements. As the number of parallel pairs increases the width of this band increases.

A set of states of the lattice of spins in case A which are susceptible to exact treatment on this model consists of the states derived from the all-parallel state by a uniform twisting of the ion spins through an angle φ per atom layer (i.e., a "spin wave" of wave vector $2N\varphi/a$). Each of these states will produce a band of electronic states the width of which is decreased relative to the aligned state by a factor $\cos\varphi/2$ due to the reduced transfer integral. Thus our model obeys Herring's theorem[8] that the energy of a spin wave is proportional to k^2 for small k. The directions of the electron spins will follow accurately those of the ions.

Case B, $J \ll b$, is essentially that envisaged by Zener[9] in his model of ferromagnetism in metals. In this case the free electrons will travel some distance while maintaining their spins without regard to the directions of the ion core spins. Here the spin up and spin down electrons will have opposite energies. Ferromagnetism, if it occurs, will do so by the mutual polarization by the ions of the free electrons, which in turn polarize the ions. The fact that this is a two-stage process again makes very reasonable the idea that χ *vs* T at high temperatures will not contain the θ term.[10]

Thus in both cases A and B what little we can say of the probable state of affairs tends to indicate that the double exchange mechanism of spin interaction is quite different qualitatively from ordinary exchange, and in particular leads to a very different form (16) for the high-temperature susceptibility, which cannot be expected to extrapolate linearly to an intercept near the Curie point. As mentioned in the abstract, this result does not fit the experimental data on the manganites.[3]

ACKNOWLEDGMENTS

We should like to acknowledge helpful discussions with M. Shimizu, K. Yosida, D. Kleinman, and H. B. Callen.

APPENDIX I. TRANSFER PROCESSES

Throughout these appendices we shall use Serber's method[6,11] of treating configuration interactions. This method, based on Dirac's method for treating spin interactions within configurations, gives us the following rules for writing down the Hamiltonian matrix:

(1) Assign orbitals to the electrons in each configuration in some fixed, convenient order. Two orbitals, treated as if distinct, are to be assigned to electron pairs which are in the same orbital.

(2) Matrix elements of the Hamiltonian within configurations are to be computed as in Dirac's method, multiplying ordinary integrals by unity and exchange integrals by the spin permutation operator $-P_{12}$ of the spins in the appropriate orbitals.

(3) Matrix elements connecting different configurations are written down by exactly the same rules in the sense that the choice between "direct" and "exchange-type" integrals is made on the basis of the order number. That is, if the two configurations are, say, $\varphi(1)\chi(2)$ and $\varphi(1)\psi(2)$ the integral $\int\chi(1)H\psi(1)d\tau$ is a "direct" type, $\int\varphi(1)\chi(2)H\varphi(2)\psi(1)d\tau$ is "exchange" and is to be multiplied by the permutation operator $-P_{12}$ in the off-diagonal block.

(4) Where one configuration has a pair of electrons in the same orbital, the other not, the transition matrix elements leading to forbidden spin arrangements are to be eliminated by appropriate projection operators, while certain factors of $\sqrt{2}$ are to be multiplied into the remaining off-diagonal blocks.

The numbering of wave functions which we shall use is shown in the following table for the three configurations of Table I. We leave out any electrons in the core states d_1' and d_2' because these are irrelevant to the transfer mechanism, and will appear only in the final problem of Appendix II.

Configuration	d_1	p	p	d_2
I	1	2	3	
II		1	3	2
III	1	3		2

The two transfer mechanisms we consider are, first, the exchange-type integral:

$$J^* = \int d_1(1)p(2)Hd_2(2)p(1)d\tau;$$

and second, the second-order effect of configuration III on configurations I and II, which will be proportional to

$$|H_{\text{III}\to d}|^2/(E_{\text{III}}-E_{\text{I}}).$$

These seem to be the only effects of sufficiently low order to be appreciable. We wish to show that each of these effects connects only the following wave functions in configurations I and II:

$$\alpha(1)[\alpha(2)\beta(3)-\alpha(3)\beta(2)]\leftrightarrow [\alpha(1)\beta(3)-\alpha(3)\beta(1)]\alpha(2), \quad \text{(A1)}$$

and similarly for the β's.

[8] C. Herring and C. Kittel, Phys. Rev. **81**, 869 (1951).
[9] C. Zener, Phys. Rev. **81**, 440 (1951).
[10] Unless the electron gas is degenerate. For a degenerate electron gas ($kT \ll b$) we can draw no conclusions for many reasons (electron interactions, second-order effects); thus the present considerations do not affect Zener's theory of ferromagnetism in metals.[9]
[11] P. W. Anderson, Phys. Rev. **79**, 350 (1950).

In the transfer "block" of the Hamiltonian connecting I and II, according to rule 3 J^* will be multiplied by P_{12}, the permutation operator connecting 1 and 2. It must also be multiplied by appropriate projection operators in order to "communicate" with configurations I and II, according to rule 4.

Transfer operator=

$$J^* O_{23} P_{12} O_{13}, \tag{A2}$$

where O_{13} and O_{23} are the projection operators eliminating triplet states for the pairs 13 or 23. First, since O_{23}, O_{13}, and P_{12} are all scalars, we know that indeed (A2) can only connect $S_z=+\frac{1}{2}$ with $+\frac{1}{2}$ and $-\frac{1}{2}$ with $-\frac{1}{2}$, so that our requirement is satisfied since (A1) are the only pair of wave functions of I and II with the right S_z's to be connected together. To get the correct numerical factor, we apply (A2) to

$$\frac{(\alpha_1\beta_2-\alpha_2\beta_1)\alpha_2}{\sqrt{2}}$$

$$J^* O_{23} P_{12} O_{13}\left(\frac{(\alpha_1\beta_2-\alpha_2\beta_1)\alpha_2}{\sqrt{2}}\right)$$

$$= J^* O_{23} P_{12}\left(\frac{(\alpha_1\beta_2-\alpha_2\beta_1)\alpha_2}{\sqrt{2}}\right)$$

$$= \frac{J^*}{\sqrt{2}} O_{23}(\alpha_1\alpha_2\beta_2-\alpha_2\alpha_1\beta_2)$$

$$= J^* \frac{\alpha_1(\alpha_2\beta_2-\alpha_2\beta_2)}{\sqrt{2}}.$$

Thus we have the expected result, that the effect of J^* is just to replace an up spin in d_1 by an up spin in d_2 and vice versa.

The problem for $|H_{pd}|^2/(E_{III}-E_I)$ is only slightly more complicated. In the perturbation procedure we go once from I to III, requiring an operator O_{23} and a factor $\sqrt{2}$, do nothing in III and go down to II with $\sqrt{2}\,O_{13}$. Thus

$$\text{Transfer operator} = -2|H_{pd}|^2 O_{23} O_{13}/(E_{III}-E_I). \tag{A3}$$

Again we note that the scalarity of O_{23} and O_{13} proves that we have a simple transfer operator, and calculate its magnitude:

$$O_{23} O_{13} \frac{(\alpha_1\beta_2-\alpha_2\beta_1)\alpha_2}{\sqrt{2}}$$

$$= O_{23} \frac{(\alpha_1\beta_2-\alpha_2\beta_1)\alpha_2}{\sqrt{2}}$$

$$= \frac{1-P_{23}}{2}\left(\frac{\alpha_1\beta_2-\alpha_2\beta_1}{\sqrt{2}}\right)\alpha_2$$

$$= \frac{1}{2\sqrt{2}}(\alpha_1\alpha_2\beta_2-\beta_1\alpha_2\alpha_2-\alpha_1\beta_2\alpha_2+\beta_1\alpha_2\alpha_2)$$

$$= -\tfrac{1}{2}(\alpha_1)\left(\frac{\alpha_2\beta_2-\alpha_2\beta_2}{\sqrt{2}}\right),$$

so that again the factor multiplying the simple transfer integral is just unity.

APPENDIX II. QUANTUM CALCULATION OF DOUBLE EXCHANGE

With the assumptions stated in the text we can write down the following Hamiltonian matrix for our problem. We take the odd electron's spin as $\mathbf{s}$, the spins of the ion cores as $\mathbf{S}_1$ and $\mathbf{S}_2$. b is the transfer integral, either J^* for nonorthogonal wave functions or $|H_{pd}|^2/(E_{III}-E_I)$ if we consider the transfer to go through configuration III.

$$H = \begin{array}{c} \\ \text{I} \\ \text{II} \end{array} \begin{array}{c} \begin{array}{cc} \text{I} & \text{II} \end{array} \\ \begin{pmatrix} -2J\mathbf{S}_1\cdot\mathbf{s} & b\mathbf{1} \\ b\mathbf{1} & -2J\mathbf{S}_2\cdot\mathbf{s} \end{pmatrix} \end{array}. \tag{A4}$$

$\mathbf{1}$ is the unit operator. As in the text, we can diagonalize the blocks I−I and II−II by choice of the appropriate representation, but only of a different one in each block: namely, in I we combine

$$\mathbf{S}_1+\mathbf{s}=\mathbf{S}_1', \quad \mathbf{S}_1'+\mathbf{S}_2=\mathbf{S}_0, \tag{A5}$$

and use the representation in which S_0, M_{S_0}, and $S_1'=S\pm\frac{1}{2}$ are diagonal. Then $-2J(s\cdot S_1)=-JS$ or $J(S+1)$ for $S_1'=S+\frac{1}{2}$ or $S-\frac{1}{2}$. Similarly, we can diagonalize II but only by setting $S_2'=S\pm\frac{1}{2}$ diagonal, in which case S_1' is not.

One of two possible ways to proceed is to use the correct coordinates to diagonalize each diagonal block and transform $b\mathbf{1}$ to connect the blocks properly. This is easily done by replacing $\mathbf{1}$ by the unitary matrix which effects the transformation between the two representations. This matrix can be obtained by the use of Racah's methods.[12,13] In the notation of reference 13, the matrix element connecting the wave function for a particular value of S_1' with that for a value of S_2' is

$$\langle S_1, sS_2(S_2'), S_0|S_1s(S_1'), S_2, S_0\rangle = [(2S_1'+1)(2S_2'+1)]^{\frac{1}{2}} W(S_1sS_0S_2; S_1'S_2'),$$

which can be easily evaluated by the tables in reference

[12] G. Racah, Phys. Rev. **62**, 438 (1943).
[13] Biedenharn, Blatt, and Rose, Revs. Modern Phys. **24**, 249 (1952).

13, using $S_1=S_2=S$; $s=\frac{1}{2}$. The resulting Hamiltonian matrix is

$$H=\begin{array}{c} \\ \text{I:} \\ \\ \\ \text{II:} \\ \\ \end{array}\begin{array}{c} \\ S_1'=S+\frac{1}{2} \\ S-\frac{1}{2} \\ S_2'=S+\frac{1}{2} \\ S-\frac{1}{2} \end{array}\begin{array}{c} \begin{array}{cccc} S_1'-S+\frac{1}{2} & S_1'=S-\frac{1}{2} & S_2'=S+\frac{1}{2} & S_2'=S-\frac{1}{2} \end{array} \\ \begin{bmatrix} -JS & 0 & (-1)^{1-2S_0}\dfrac{S_0+\frac{1}{2}}{2S+1} & \left[1-\left(\dfrac{S_0+\frac{1}{2}}{2S+1}\right)^2\right]^{\frac{1}{2}} \\ 0 & J(S+1) & -\left[1-\left(\dfrac{S_0+\frac{1}{2}}{2S+1}\right)^2\right]^{\frac{1}{2}} & (-1)^{1-2S_0}\dfrac{S_0+\frac{1}{2}}{2S+1} \\ (-1)^{1-2S_0}\dfrac{S_0+\frac{1}{2}}{2S+1} & -\left[1-\left(\dfrac{S_0+\frac{1}{2}}{2S+1}\right)^2\right]^{\frac{1}{2}} & -JS & 0 \\ \left[1-\left(\dfrac{S_0+\frac{1}{2}}{2S+1}\right)^2\right]^{\frac{1}{2}} & (-1)^{1-2S_0}\dfrac{S_0+\frac{1}{2}}{2S+1} & 0 & J(S+1) \end{bmatrix} \end{array}. \quad \text{(A6)}$$

This is identical with the Hamiltonian for the simple model, Eq. (7), if we set $\cos\theta/2=(S_0+\frac{1}{2})/(2S+1)$ as in Eq. (10).

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Effect of Pressure on the Electrical Resistance of Metals at Liquid Helium Temperatures

J. HATTON
Department of Physics, Harvard University, Cambridge, Massachusetts
(Received June 17, 1955)

The effect of pressures up to 5000 kg/cm² on the electrical resistance of polycrystalline copper, silver, gold, and platinum and on single crystals of arsenic, antimony, and bismuth has been measured at liquid helium temperatures. The pressures were generated by a piston and cylinder arrangement using solid hydrogen as the pressure transmitting medium.

INTRODUCTION

IT is generally accepted that the electron-lattice interaction in a metal at low temperatures is an important factor in determining whether the metal will have a superconducting phase. An obvious method of seeking information about this interaction is to investigate how the electrical conductivity of metals at low temperatures is influenced when the lattice spacings are changed by application of pressure. In the case of nonsuperconductors, such measurements should be of intrinsic interest in the theory of metallic conduction, since at liquid helium temperatures the contribution to resistance from thermal scattering is small and the change with pressure from this cause may be expected to be negligible.

EXPERIMENTAL

A major technical problem to be faced is the method of applying hydrostatic pressure to the metallic specimen. Since even helium solidifies at quite low pressure in this temperature region, one is compelled to use a solid for transmission of the pressure, and if the stress transmitted to the specimen is to approximate a hydrostatic pressure, then a transmitting medium which will not support any appreciable shear stress is required. Presumably, helium would be the most satisfactory substance from this standpoint, and it might be practicable to use helium in a bomb type of apparatus similar to that which has been used in studying the thermodynamic properties of helium under pressure.[1] However,

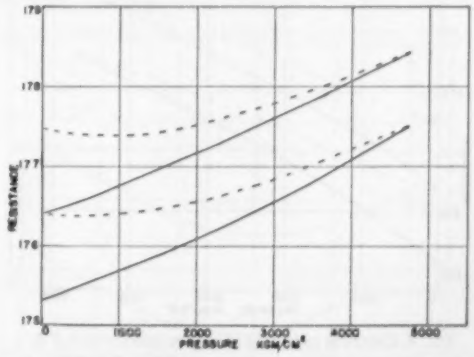


FIG. 1. Electrical resistance of copper *vs* pressure at 4.2°K.

[1] J. S. Dugdale and F. E. Simon, Proc. Roy. Soc. (London) **A218**, 291 (1953).

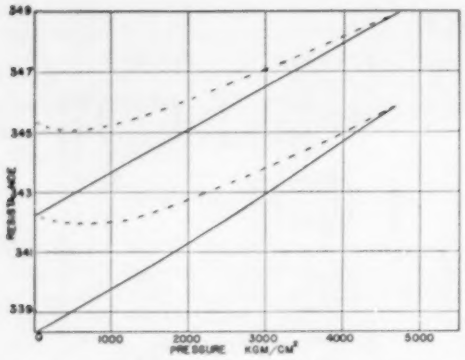


FIG. 2. Electrical resistance of silver *vs* pressure at 4.2°K.

if one wishes to use a piston and cylinder arrangement for applying the pressure, there is a difficult problem of sealing the piston against liquid helium which does not solidify until pressure is applied.

For an exploratory investigation, it was decided to use a piston and cylinder arrangement with solid hydrogen as the pressure-transmitting medium. The work of Stewart and Swenson[2,3] showed that solid hydrogen at liquid helium temperatures flows readily and extrudes through small holes at quite low pressures, suggesting that it might be suitable for transmitting an approximately hydrostatic pressure.

An apparatus similar to that used by the above workers was constructed, with the necessary additions for making electrical measurements. Four electrical leads entered the pressure cylinder via its closed end through pipestone seals of the type described by Bridgman,[4] and the resistance of the specimens was determined from current and potential measurements by means of a potentiometer circuit.

Each sample to be measured had four short copper leads soldered to it, using ordinary soft solder in the case of copper, silver, and platinum, but using a low melting point solder for gold, arsenic, antimony, and bismuth. Connections were then made to the insulated leads inside the pressure chamber by ordinary soft solder.

The current through the specimen was adjusted to give a potential drop of the order of 100μ volt and had values within the range 0.25–1.25 amp. During measurements, the pressure cylinder was completely immersed in liquid helium whose temperature could be reduced to about 1.6°K.

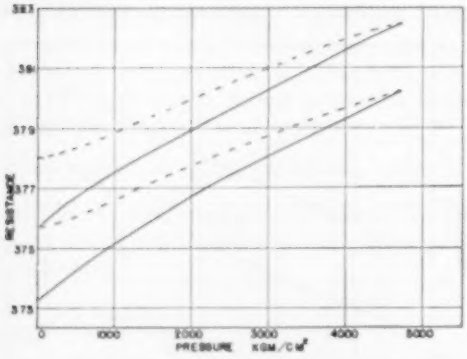
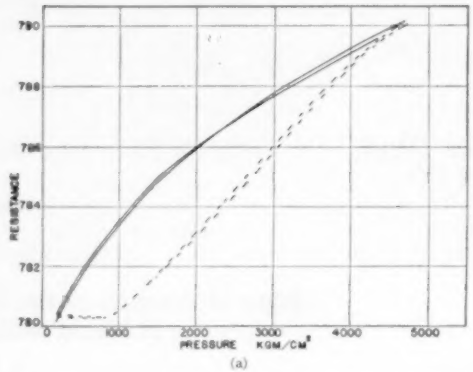


FIG. 3. Electrical resistance of gold *vs* pressure at 4.2°K.



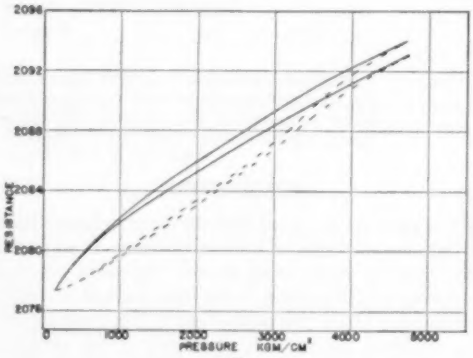


FIG. 4. (a) Electrical resistance of platinum (annealed) *vs* pressure at 4.2°K. (b) Electrical resistance of platinum (unannealed) *vs* pressure at 4.2°K.

RESULTS

The results to be presented refer to the actually measured resistances (in arbitrary units) of the specimens and not to their specific resistances. The graphs show resistance at 4.2°K as a function of pressure for two complete pressure cycles after the behavior had become reasonably reproducible as a consequence of the full pressure having been applied and released a few times. In all the figures, the full lines refer to measure-

[2] J. W. Stewart and C. A. Swenson, Phys. Rev. **94**, 1069 (1954).
[3] J. W. Stewart, Phys. Rev. **97**, 578 (1955).
[4] P. W. Bridgman, Proc. Am. Acad. Sci. **74**, No. 2, 11–20 (1940).

ments made with pressure increasing and the broken lines to measurements with pressure decreasing.

To give some idea of the reproducibility from sample to sample, we show in Figs. 4(a) and 4(b) the results for two different specimens of platinum; these results are unusual in that the curves corresponding to pressure decreasing lie below the curves corresponding to pressure increasing.

An indication of the precision of the measurements is provided by the experimental points for bismuth shown in Fig. 10.

(*i*) *Copper.*—A piece of stout copper wire of high purity (99.99%) was kindly made available to me by Dr. C. A. Swenson of Massachusetts Institute of Technology. This was rolled to a foil from which a thin ribbon was cut and roughly vacuum annealed before assembly in the pressure chamber. The results are shown in Fig. 1.

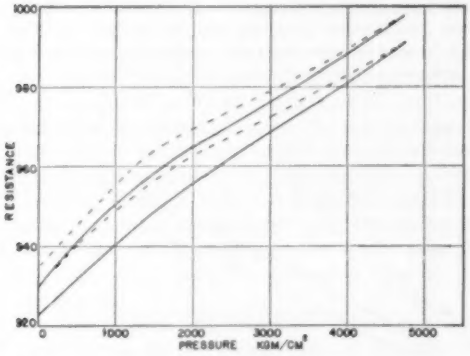


FIG. 5. Electrical resistance in the plane of easy cleavage of crystalline arsenic *vs* pressure at 4.2°K.

(*ii*) *Silver.*—The specimen was a 0.003-in. diam wire of "fine silver" of unspecified purity. Before mounting in the pressure chamber it was roughly annealed in vacuum. The results are shown in Fig. 2.

(*iii*) *Gold.*—The specimen was in the form of a thin ribbon which had been prepared by rolling down heavy gold wire of unspecified purity. It was crudely annealed in vacuum before assembly in the pressure chamber. The results are shown in Fig. 3.

(*iv*) *Platinum.*—The platinum was in the form of wire of unspecified purity. Figure 4 shows the results for a specimen of 0.002-in. diameter which had been roughly vacuum annealed. For comparison, the results for a 0.003-in. diameter wire which had not been annealed are shown in Fig. 4(b).

(*v*) *Arsenic.*—Some single-crystal arsenic of high purity was kindly made available to me by Professor Bridgman. For measurements of the resistance parallel to the planes of easy cleavage it was relatively simple to cleave off a flake of thickness ~0.1 mm, from which a

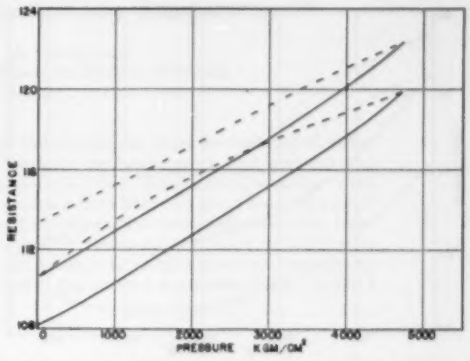


FIG. 6. Electrical resistance perpendicular to plane of easy cleavage of crystalline arsenic *vs* pressure at 4.2°K.

specimen of length ~4 mm and width ~0.5 mm was prepared.

However, for resistance measurements perpendicular to the planes of easy cleavage it was necessary to cut from a crystal a needle-shaped specimen with the cleavage planes perpendicular to its length. This proved to be exceedingly difficult on account of the extreme fragility of the arsenic, and the best specimens produced were about 1 mm long and 0.5 mm in transverse dimensions.

The results at 4.2°K for the two crystal directions are shown in Figs. 5 and 6. There was no significant difference when the temperature was reduced to 1.7°K.

It is interesting to observe that the arsenic specimens, in spite of their fragility, were undamaged by application of the high pressure.

(*vi*) *Antimony.*—The high purity, single crystal antimony was obtained from Professor Bridgman. From a thin cleaved flake specimens of length ~5 mm and breadth ~0.3 mm were prepared, while for measurements perpendicular to the planes of easy cleavage specimens about 4 mm long and 0.8 mm in transverse dimensions were cut.

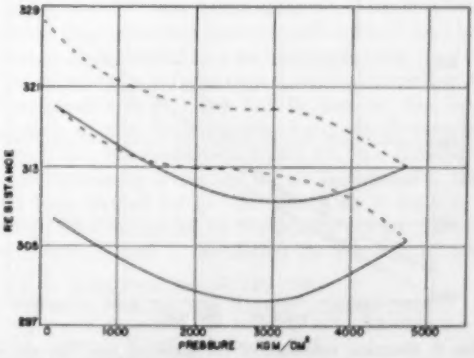


FIG. 7. Electrical resistance in the plane of easy cleavage of crystalline antimony *vs* pressure at 4.2°K.

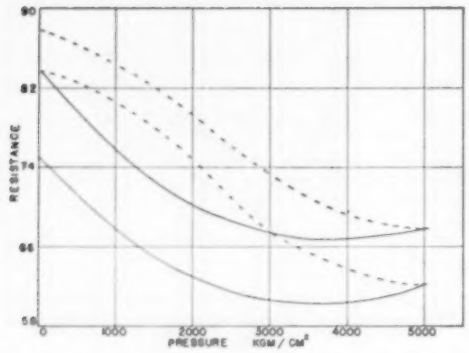


FIG. 8. Electrical resistance perpendicular to plane of easy cleavage of crystalline antimony *vs* pressure at 4.2°K.

The results at 4.2°K are shown in Figs. 7 and 8. In neither crystal direction was there any significant change of behavior when the temperature was reduced to ~2°K.

(*vii*) *Bismuth.*—Specimens of similar dimensions to those used in the case of antimony were secured from single crystals of pure bismuth which also came from Professor Bridgman.

The results are shown in Figs. 9 and 10. The resistance perpendicular to the planes of easy cleavage exhibits a rather unusual behavior. With increasing pressure there is a shallow minimum at about 250 kg/cm² followed by a practically linear increase. When the pressure is reduced from its highest value the resistance rises fairly rapidly to a maximum at about 3200 kg/cm² and then decreases steadily as the pressure is further reduced. The minimum in the neighborhood of 250 kg/cm² was checked by a number of measurements in which the pressure was increased and decreased by small steps within the range from zero to 400 kg/cm².

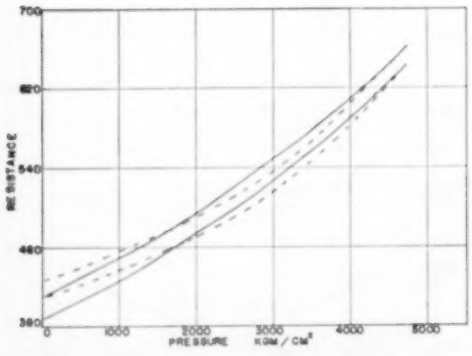


FIG. 9. Electrical resistance in the plane of easy cleavage of crystalline bismuth *vs* pressure at 4.2°K.

The behavior in both crystal directions was essentially the same when the temperature was reduced to ~1.7°K.

DISCUSSION

The results presented in the preceding section were checked in each case by at least two separate sets of measurements on different specimens. Although the qualitative agreement was good, the quantitative results for the fractional change of resistance with pressure were not found to be reproducible to better than a factor of about 2. As yet, insufficient investigations have been made to know the extent to which these discrepancies are influenced by such things as the physical and chemical purity of the metal or by the degree of nonuniformity of the stress transmitted by solid hydrogen. Here we might point out that the fact that the arsenic specimens could be subjected to many full excursions of the pressure without damage, in spite of their extreme fragility, suggests that they could not have been exposed to any very large shearing forces.

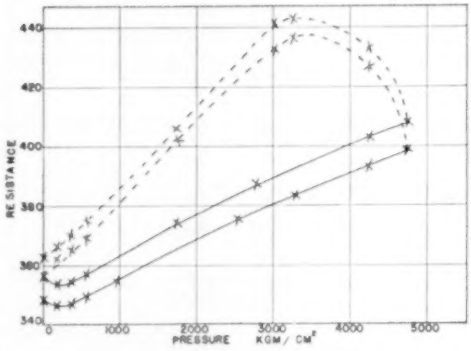


FIG. 10. Electrical resistance perpendicular to plane of easy cleavage of crystalline bismuth *vs* pressure at 4.2°K.

It is evident from the figures that after each cycle of pressure change the resistance does not return to its initial value. This might be interpreted as evidence that the metal is being worked as a result of nonuniform stresses. It seems that this question will be solved only by further investigations using other means for transmitting the pressure, and work along these lines is in progress.

In conclusion, we repeat that the experiments reported here are only of a preliminary nature, but there can be little doubt that they give a fair indication of the variation of the low-temperature resistance with hydrostatic pressure. The relatively simple techniques employed lead to the possibility of a wide field of electrical measurements under high pressure at liquid helium temperatures. There is no reason why the obtainable pressures should not be extended considerably above those used here.

Ferromagnetic and Antiferromagnetic Curie Temperatures*†

H. A. BROWN‡ AND J. M. LUTTINGER§
Department of Physics, University of Wisconsin, Madison, Wisconsin

(Received July 7, 1955)

The methods of calculating Curie temperatures used by Opechowski and Weiss are extended to higher spin values. Using the former, Curie temperatures are found for the face-centered, body-centered, and simple cubic and the hexagonal and quadratic layer lattices for spin $\frac{1}{2}$ to 3. Weiss' method is applied to the body-centered and simple cubics for spin $\frac{1}{2}$ to 2 and to the quadratic layer lattice for spin $\frac{1}{2}$ and 1. Numerical results show good agreement between the two methods for three-dimensional ferromagnetic lattices. However, the Opechowski method fails to locate the antiferromagnetic transition and predicts a transition for the two-dimensional lattices while the Weiss method does not. A classical-spin approximation is investigated and is found to give Curie temperatures within a few percent of the quantum-mechanical results for spin 1 or greater.

I. INTRODUCTION

THE Heisenberg model of a ferromagnetic[1] provides a means of calculating Curie temperatures which has the advantage of involving only one parameter, the exchange integral, while retaining some of the physical properties of the system. The present work is concerned with the extension of two previous calculations, both based on the Heisenberg model, to the values of spin greater than $\frac{1}{2}$. These two methods begin with the same Hamiltonian and arrive at predictions of the Curie temperatures by alternate routes; the one proceeding, without further assumption, to find the high-temperature series expansion of the partition function, whereas the other, by means of a physical approximation, treats a simplified problem in semiclosed form. The agreement of the results of the two methods would seem to indicate that they are both reasonably close to the exact solution.

The first method was suggested by Kramers[2] and carried through by Opechowski[3] and is hereafter referred to as the K-O method. We begin with the Heisenberg Hamiltonian for a lattice of N atoms with nearest-neighbor (n.n.) exchange:

$$\mathcal{H}=-2J\sum \mathbf{S}_i\cdot\mathbf{S}_j-g\beta\mathbf{H}_0\cdot\sum \mathbf{S}_k, \qquad (1)$$

where $\mathbf{S}_i$ is the spin operator for the ith atom, J the exchange integral, and $\mathbf{H}_0$ the external field. The first sum is over all pairs of n.n. atoms and the second over all atoms. The partition function,

$$Z=\text{trace}\exp(-\mathcal{H}/kT), \qquad (2)$$

is evaluated by expanding the exponential and evaluating traces of successively higher powers of $(\mathcal{H}/kT)$. The Curie temperature, which will manifest itself as a singularity in the partition-function or one of its derivatives, is inferred from the behavior of the power series expansions, which, for practical reasons, are limited to four or five terms.

The second calculation was performed by Weiss[4] using a technique due to Bethe[5] and modified by Peierls.[6] The Bethe-Peierls-Weiss (B-P-W) method replaces the Heisenberg Hamiltonian for the entire crystal by the corresponding Hamiltonian for a cluster formed by an atom and its n.ns. Using the simplifying assumption that the interactions of the cluster with the rest of the crystal can be replaced by an internal field, $\mathbf{H}_1$, the cluster Hamiltonian is then

$$\mathcal{H}_{cl}=-2J\mathbf{S}_0\cdot\sum \mathbf{S}_i-g\beta\mathbf{H}_1\cdot\sum \mathbf{S}_i-g\beta\mathbf{H}_0\cdot\mathbf{S}_0, \qquad (3)$$

where $\mathbf{S}_0$ is the spin of the central atom which has n n.ns. with spins $\mathbf{S}_i$ $(i=1,\cdots n)$. Using $\mathcal{H}_{cl}$ instead of $\mathcal{H}$, the cluster partition function can be obtained as a power series in the fields $\mathbf{H}_0$ and $\mathbf{H}_1$, but in closed form with respect to the exchange energy. The Curie temperature is located by the existence of a spontaneous magnetization, the internal field being eliminated by a consistency condition.

Both of the above methods have been somewhat limited in their application. The K-O method has been carried to the fourth stage in the approximation (i.e., the traces of $\mathcal{H}^n$ were found up to $n=4$) for the close-packed hexagonal and face-centered cubic (f.c.c.) lattices by Opechowski,[3] and for the simple cubic (s.c.) by Zehler,[7] who also includes some numerical corrections to Opechowski's work. The B-P-W method has been applied to the ferromagnetic b.c.c. (body-centered cubic), s.c., h.l. (hexagonal layer), and q.l. (quadratic layer) lattices by Weiss and the antiferromagnetic case has been treated by Li.[8] All this work is for spin $\frac{1}{2}$ except for a calculation by Weiss using the b.c.c. lattice with spin 1, and a calculation by Van Vleck,[9] who

* Based on a dissertation submitted by H. A. Brown in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Wisconsin

† This work was supported in part by the Wisconsin Alumni Research Foundation.

‡ Present address: Department of Physics, University of Miami, Coral Gables, Florida.

§ Present address: Department of Physics, University of Michigan, Ann Arbor, Michigan.

[1] W. Heisenberg, Z. Physik **49**, 619 (1928).

[2] H. A. Kramers, Leiden Supplement No. 83, 1936 (unpublished).

[3] W. Opechowski, Physica **4**, 181 (1937); **6**, 1112 (1938).

[4] P. R. Weiss, Phys. Rev. **74**, 1493 (1948).

[5] H. Bethe, Proc. Roy. Soc. (London) **A150**, 552 (1935).

[6] R. Peierls, Proc. Roy. Soc. (London) **A154**, 207 (1936).

[7] V. Zehler, Z. Naturforsch. **5a**, 344 (1950).

[8] Y. Y. Li, Phys. Rev. **84**, 721 (1951).

[9] J. H. Van Vleck, J. Chem. Phys. **6**, 105 (1938).

carried the K-O method to the second approximation for arbitrary spin.

In the present work, the K-O method will be carried to the fourth approximation and the susceptibility obtained as a function of the spin quantum number, S, for the f.c.c., b.c.c., s.c., h.l., and q.l. lattices (Part II). Numerical solutions for the Curie temperatures are found for $S=\frac{1}{2}$ to $S=3$. The B-P-W method will be applied to the b.c.c., s.c., and q.l. (the method cannot be applied, in the present form, to lattices in which n.ns. to a given atom are n.ns. to each other) and numerical solutions found for $S=\frac{1}{2}$ to $S=2$. (Part III.) Finally, a classical-spin approximation is investigated (Part IV) which allows the extension of the B-P-W method to higher spin without the need for machine computation. A discussion of the numerical results is given in Part V.

II. KRAMERS-OPECHOWSKI METHOD

Since the procedure is adequately explained by Opechowski,[3] we will give only a brief outline. By the rules of statistical mechanics, the zero-field susceptibility per atom is

$$\chi=\frac{kT}{N}\left(\frac{\partial^2 \ln Z}{\partial H_0^2}\right)_{H_0=0}=\frac{g^2\beta^2}{NkT}\left(\frac{\partial^2 \ln Z}{\partial \lambda_0^2}\right)_0, \tag{4}$$

with

$$\lambda_0=\frac{g\beta H_0}{kT}.$$

This gives us the expression

$$\chi=\frac{g^2\beta^2 S(S+1)}{3kT}\sum_{n=0}^{\infty}\frac{a_n}{t^n}, \tag{5}$$

where $t=kT/J$, and the a_n are lattice and spin-dependent coefficients which are related to the A_n of Opechowski[3] by

$$a_n=-\frac{3}{S(S+1)}\left(\frac{d^2A_n}{d\lambda_0^2}\right)_0. \tag{6}$$

The evaluation of the susceptibility is thus formally reduced to the problem of evaluating the coefficients a_n, which can be shown to be equal to a sum of products of numerical factors which depend on the number of ways certain "graphs" can be drawn on the lattice, and traces which depend only on the spin and can be found directly as functions of S. Details of this calculation will be published later. The results are:

$$a_0=1, \tag{7}$$

$$a_1=(4/3)f\bar{S}^2, \tag{8}$$

$$a_2=(8/9)f(2f-1)\bar{S}^4-\tfrac{2}{3}f\bar{S}^2, \tag{9}$$

$$a_3=(8/27)f\bar{S}^2\{[2(2f-1)^2-2L_2-(6/5)]\bar{S}^4 -[3(2f-1)+L_2+(6/5)]\bar{S}^2+(6/5)\}, \tag{10}$$

$$a_4=(f\bar{S}^2/1215)\{A\bar{S}^6-B\bar{S}^4+C\bar{S}^2-270\}, \tag{11}$$

where $f=\frac{1}{2}n$=one-half the number of n.ns. and $\bar{S}^2$ means $S(S+1)$, and

$$A=\begin{vmatrix}578208\\154848\\55200\\44832\\10272\end{vmatrix},\quad B=\begin{vmatrix}158592\\57792\\29280\\33168\\10848\end{vmatrix},\quad C=\begin{vmatrix}14112\\8352\\6120\\6768\\3888\end{vmatrix};\quad L_2=\begin{vmatrix}4\\0\\0\\2\\0\end{vmatrix}.$$

The "column vectors" of numbers refer to the five lattices considered, the convention being that from top to bottom, the numbers refer to the f.c.c., b.c.c., s.c., h.l., and q.l. lattices, respectively.

Denoting by χ_n the nth approximation to the susceptibility, we see that the zeroth approximation is Weiss' law $\chi_0=g^2\beta^2\bar{S}^2/3kT$, and the first approximation is the Curie-Weiss law with $kT_c=(4/3)JfS(S+1)$, which is the molecular field theory result.[10]

The susceptibility of a ferromagnetic has a singularity at the Curie temperature which will appear as a root of the inverse susceptibility. We attempt to locate this singularity by finding the roots of the successive approximations, $\chi_n^{-1}=0$, obtaining the numbers that are found in Table I under the heading, "K-O Method, QM, roots." For the second approximation, setting $\chi_2^{-1}=0$ gives a quadratic equation of which the high-temperature root is

$$\frac{J}{kT_c}=\frac{1}{4\bar{S}^2+1}\left\{1-\left(1-\frac{4\bar{S}^2+1}{2f\bar{S}^2}\right)^{\frac{1}{2}}\right\}. \tag{12}$$

For $S=\frac{1}{2}$, this reduces to the result found by Heisenberg,[1] and for $S=1$ to that found by Nakano[11] who duplicated the Heisenberg calculation for spin 1.

An alternate procedure for finding the singularity of a function expressed as a power series is to find the radius of convergence of the series. Thus, if $f=\sum_n c_n x^n$, the radius of convergence is $x_c=\lim c_{n-1}/c_n$ for $n\to\infty$. This method, applied to $\sum_n a_n/t^n$ gives the series of approximations $kT_c/J=a_n/a_{n-1}$ which are listed in Table I under "K-O Method, QM, ratios."

It is a weakness of the K-O method that there is not a unique criterion for defining the Curie temperature. An inspection of the results of the two methods employed here shows that successive approximations always seem to be converging to some definite T_c, that the rapidity of convergence is such that the fourth approximation is usually enough to locate the convergence limit within several percent, and that the limits of convergence of the two methods are the same or nearly the same in all cases. There is therefore, *a posteriori*, justification of the series expansion method and of the criteria used to define the Curie temperatures for ferromagnetics.

For antiferromagnetics, the only change is in the

[10] J. H. Van Vleck, J. Chem. Phys. **9**, 85 (1941).
[11] H. Nakano, Progr. Theoret. Phys. (Japan) **9**, 403 (1953).

TABLE I. Calculated values of (kT_c/J).

Spin			Kramers-Opechowski method f.c.c. $n=1$	$n=2$	$n=3$	$n=4$	b.c.c. $n=1$	$n=2$	$n=3$	$n=4$	s.c. $n=1$	$n=2$	$n=3$	$n=4$	h.l. $n=1$	$n=2$	$n=3$	$n=4$	q.l. $n=1$	$n=2$	$n=3$	$n=4$
1/2	*QM*	Roots	6.00	4.73	4.26	4.24	4.00	2.00	2.95	2.39	3.00	(1.73)	2.00	1.93	3.00	(1.73)	(1.86)	1.75	2.00	(1.41)	1.25	1.09
		Ratios	6.00	5.00	4.60	4.43	4.00	3.00	2.89	2.76	3.00	2.00	1.83	1.88	3.00	2.00	1.42	1.10	2.00	1.00	0.67	0.81
	Class	Roots	6.00	5.45	5.23	5.11	4.00	3.41	3.45	3.25	3.00	2.37	2.43	2.28	3.00	2.37	1.76	(1.81)	2.00	1.00	1.38	(1.22)
		Ratios	6.00	5.50	5.29	5.17	4.00	3.50	3.46	3.33	3.00	2.50	2.44	2.36	3.00	2.50	2.24	2.08	2.00	1.50	1.40	1.27
1	*QM*	Roots	16.00	13.89	13.10	12.74	10.67	8.31	8.66	7.82	8.00	5.16	5.93	5.36	8.00	5.16	(4.79)	(4.53)	5.33	(3.13)	3.21	(3.01)
		Ratios	16.00	14.17	13.43	13.04	10.67	8.83	8.65	8.27	8.00	6.17	5.91	5.70	8.00	6.17	5.19	4.62	5.33	3.50	3.05	2.68
	Class	Roots	16.00	14.53	13.94	13.63	10.67	9.10	9.21	8.67	8.00	6.31	6.49	6.08	8.00	6.31	4.70	(4.82)	5.33	2.67	3.68	(3.26)
		Ratios	16.00	14.67	14.11	13.80	10.67	9.53	9.22	8.89	8.00	6.67	6.51	6.28	8.00	6.67	5.97	5.56	5.33	4.00	3.73	3.40
3/2	*QM*	Roots	30.00	26.62	25.32	24.67	20.00	16.33	16.71	15.42	15.00	10.86	11.60	10.63	15.00	10.86	(8.80)	(8.94)	10.00	(5.48)	6.32	(5.94)
		Ratios	30.00	27.00	25.78	25.11	20.00	17.00	16.72	16.03	15.00	12.00	11.61	11.18	15.00	12.00	10.42	9.49	10.00	7.00	6.32	5.63
	Class	Roots	30.00	27.25	26.14	25.56	20.00	17.07	17.27	16.25	15.00	11.83	12.17	11.40	15.00	11.83	8.82	(9.04)	10.00	5.00	6.90	(6.10)
		Ratios	30.00	27.50	26.45	25.87	20.00	17.50	17.29	16.67	15.00	12.50	12.20	11.78	15.00	12.50	11.20	10.42	10.00	7.50	7.00	6.37
2	*QM*	Roots	48.00	42.98	41.02	40.00	32.00	26.59	27.08	25.17	24.00	18.00	18.89	17.46	24.00	18.00	(13.93)	(14.28)	16.00	(8.48)	10.43	(9.63)
		Ratios	48.00	43.50	41.65	40.64	32.00	27.50	27.09	26.03	24.00	19.50	18.93	18.24	24.00	19.50	17.15	15.76	16.00	11.50	10.53	9.44
	Class	Roots	48.00	43.60	41.83	40.89	32.00	27.31	27.64	26.00	24.00	18.93	19.48	18.23	24.00	18.93	14.11	(14.46)	16.00	8.00	11.05	(9.77)
		Ratios	48.00	44.00	42.33	41.40	32.00	28.00	27.66	26.66	24.00	20.00	19.52	18.85	24.00	20.00	17.92	16.68	16.00	12.00	11.20	10.19
5/2	*QM*	Roots	70.00	62.95	60.19	58.74	46.67	39.11	39.75	37.10	35.00	26.70	27.82	25.81	35.00	26.70	(20.19)	(20.76)	23.33	(12.16)	15.48	(14.13)
		Ratios	70.00	63.67	61.05	59.61	46.67	40.33	39.77	38.25	35.00	28.67	27.87	26.88	35.00	28.67	25.36	23.39	23.33	17.00	15.66	14.11
	Class	Roots	70.00	63.58	61.00	59.63	46.67	39.83	40.30	37.91	35.00	27.60	28.40	26.59	35.00	27.60	20.58	(21.09)	23.33	11.67	16.11	(14.24)
		Ratios	70.00	64.17	61.73	60.37	46.67	40.83	40.33	38.88	35.00	29.17	28.47	27.49	35.00	29.17	26.13	24.32	23.33	17.50	16.33	14.86
3	*QM*	Roots	96.00	86.56	82.85	80.89	64.00	53.91	54.70	51.19	48.00	36.97	38.37	35.74	48.00	36.97	(27.58)	(28.77)	32.00	(16.49)	21.45	(19.43)
		Ratios	96.00	87.50	83.98	82.03	64.00	55.50	54.75	52.69	48.00	39.50	38.45	37.09	48.00	39.50	35.07	32.43	32.00	23.50	21.73	19.62
	Class	Roots	96.00	87.19	83.66	81.77	64.00	54.63	55.27	52.00	48.00	37.86	38.95	36.47	48.00	37.86	28.22	(28.92)	32.00	16.00	22.09	(19.54)
		Ratios	96.00	88.00	84.65	82.79	64.00	56.00	55.31	53.32	48.00	40.00	39.04	37.70	48.00	40.00	35.84	33.36	32.00	24.00	22.40	20.38

Spin		Bethe-Peierls-Weiss method	b.c.c.	s.c.	q.l.
1/2	*QM*	Ferro	2.91	1.85	no
		Antiferro	3.16	2.01	root
	Class	"Exact"	3.46	2.44	1.40
		"Asymptotic"	3.41	2.37	1.00
1	*QM*	Ferro	8.72	5.91	no
		Antiferro	9.03	6.19	root
	Class	"Exact"	9.22	6.50	3.72
		"Asymptotic"	9.10	6.31	2.67
3/2	*QM*	Ferro	16.72	11.59	
		Antiferro	17.04	11.91	
	Class	"Exact"	17.28	12.19	6.97
		"Asymptotic"	17.07	11.83	5.00
2	*QM*	Ferro	27.09	18.91	
		Antiferro	27.41	19.23	
	Class	"Exact"	27.65	19.50	11.16
		"Asymptotic"	27.31	18.93	8.00
5/2	Class	"Exact"	40.32	28.45	16.28
		"Asymptotic"	39.83	27.60	11.67
3	Class	"Exact"	55.30	39.00	22.30
		"Asymptotic"	54.63	37.86	16.00

sign of J. It follows that

$$\chi=\frac{g^2\beta^2S(S+1)}{3kT}\sum_{n=0}^{\infty}(-1)^n\frac{a_n}{t^n}. \tag{13}$$

Experimental observations show that χ is finite at all temperatures but has a cusp-like maximum at the Curie temperature. Several attempts to locate this maximum using the series expansions led to inconsistent results, the convergence limits being different for each criterion used.[12] It is believed that the antiferromagnetic susceptibility may be composed of two analytic functions, the Curie temperature being merely the intersection of these two which is impossible to locate since the series expansion in $1/T$ would give information about the high-temperature function only.

III. BETHE-PEIERLS-WEISS METHOD

The cluster approach of Bethe is applicable to both ferromagnetics and antiferromagnetics. Using the cluster Hamiltonian (which is of the form of the Zeeman effect interaction with strong L-S coupling included), Weiss[4] and Li[8] find the eigenvalues exactly and evaluate the partition function as a power series in the applied and internal fields. For spin 1 and greater the diagonalization operation becomes very cumbersome and we have resorted to a technique for obtaining the trace directly as a power series in the fields.[13] This method, which is equivalent to the use of second-order perturbation theory for the eigenvalues of $\mathcal{H}_{cl}$, gives the following result for the partition function to second order in the fields (see Appendix A):

$$\begin{aligned}Z_{cl}=&\sum s_1\sum_S w(S_1)e^{p\sigma_0}(2S+1/S(S+1))\\&\times\{S(S+1)+(1/24)(\lambda_0\sigma_0+\lambda_1\sigma_1)^2\\&+(1/24)(\lambda_0-\lambda_1)^2\sigma_0\sigma_1/pS(S+1)\}\\&+\sum_S\delta_S{}^{S_0}w(S_0)\{1-4(\lambda_0-\lambda_1)^2S_0(S_0+1)/24p\}.\end{aligned} \tag{14}$$

The sum over S is from $|S_1-S_0|$ to S_1+S_0, the sum over S_1 from 0 to nS_0, excluding the term $S_1=S_0$, $S=0$, which is accounted for by the last sum in (14). The symbols are defined as follows:

$n=2f=$ number of n.ns. to a given atom,

$\mathbf{S}_0=$ spin of the central atom in the cluster,

$\mathbf{S}_1=\sum_i\mathbf{S}_i=$ resultant spin of the n.n. atoms,

$\lambda_0=g\beta\mathbf{H}_0/kT$,

$\lambda_1=g\beta\mathbf{H}_1/kT$,

$p=J/kT$,

$\sigma_1=S(S+1)+S_1(S_1+1)-S_0(S_0+1)$,

$\sigma_0=S(S+1)-S_1(S_1+1)+S_0(S_0+1)$,

$\delta_s{}^0=$ the Kronecker delta,

and $w(S_1)$ is the number of ways of forming a spin of magnitude S_1 from the n components S_i and is computed from the relation $w(S_1)=v(S_1)-v(S_1+1)$ where $v(M)$ is the coefficient of X^M in the expansion of

$$(X^{S_0}+X^{S_0-1}+X^{S_0-2}\cdots+X^{-S_0})^n.$$

The magnetic moments of the central atom and one of the n.n. atoms are, respectively,

$$m_0=g\beta(\partial\ln Z_{cl}/\partial\lambda_0), \tag{15}$$

and

$$m_1=(g\beta/n)(\partial\ln Z_{cl}/\partial\lambda_1). \tag{16}$$

The internal field, $\mathbf{H}_1$, is eliminated in principle by the requirement that, for a ferromagnetic

$$C^F: \qquad m_0=m_1, \tag{17}$$

and for an antiferromagnetic in zero external field,

$$C_0{}^{AF}: \qquad m_0=-m_1 \quad (H_0=0). \tag{18}$$

These conditions are required at all temperatures. Just below the Curie temperature, when H_1 is still small ($\lambda_1\ll1$), the partition function (14) allows us to obtain the magnetic moments to linear terms in λ_1. The "consistency" conditions (17) and (18) for zero external field then require that the temperature-dependent coefficient of λ_1 shall vanish. This leads to a complicated transcendental equation, the roots of which are found by machine. The equation found is

$$\begin{aligned}&C_0{}^{F,AF}(T_c):\\&\sum s_1\sum_S w(S_1)\exp(p_c\sigma_0)(2S+1/S(S+1))\\&\quad\times\{\sigma_0(n\sigma_1\mp\sigma_0)-(n\pm1)\sigma_0/p_cS(S+1)\}\\&\quad+\sum_S\delta_S{}^0 4(n\pm1)S_0(S_0+1)/p_c=0,\end{aligned} \tag{19}$$

the upper sign referring to the ferromagnetic and the lower to the antiferromagnetic case and the summation intervals are as above. The roots, $p_c=J/kT_c$, were obtained for $n=8$ (b.c.c.) and $n=6$ (s.c.), for $S_0=\frac{1}{2}$ to 2, and for $n=4$ (q.l.), for $S_0=\frac{1}{2}$ and 1 no roots were found. (The $S_0=\frac{1}{2}$ values agree with those previously found by Weiss and Li but a discrepancy exists between our result and that of Weiss for b.c.c. with spin 1; the values of p_c found being 0.11465 and 0.1502, respectively.)

The numerical values of kT_c/J are included in Table I under "B-P-W Method, QM, Ferro" and ". . . Antiferro." A striking feature of the results for the b.c.c. and s.c. lattices is that they agree so well with those of the third K-O approximation, especially for large spin. (The two methods are not identical since some of the terms which contribute to the third approximation involve next-nearest neighbor interactions which are not included in the cluster approach.)

It should be mentioned that although the present approach to the B-P-W method yields an equation for the Curie temperatures which is applicable to all spin values and several lattices, it cannot be used to obtain more information about the transition. This is because

[12] The methods used to define the ferromagnetic Curie temperatures give the same results here except for the sign of T_c. The negative Curie temperatures are merely the asymptotes of the $\chi-T$ curve and are not, without additional assumptions, related to the positive transition temperature.

[13] R. Karplus and J. Schwinger, Phys. Rev. 73, 1020 (1948), Appendix I.

such things as the specific heat and the spontaneous magnetization require the explicit elimination of the internal field. This means obtaining the partition function to λ_1^4 instead of λ_1^2 which, by the present method, is out of the question.

IV. CLASSICAL-SPIN APPROXIMATION

The classical-spin approximation consists of replacing the quantum-mechanical spin operators by classical vectors fixed in length, but free to orient in any direction. [In the equations following, the magnitude of the spin is written as S, but in numerical use, we replace S^2 by $S(S+1)$.] The classical partition function is an integral over phase space defined by the solid angle available to each spin vector. Thus,

$$Z=\int \exp(-\mathcal{H}/kT)d\Gamma,$$

where, in the K-O method,

$$d\Gamma=\prod_{i=1}^{N} d\Omega_i, \tag{20}$$

and $d\Omega_i$ is the element of solid angle in the direction $\mathbf{S}_i$. Although the classical approximation gives nothing that cannot be found by the quantum-mechanical K-O calculation, it does serve as a useful check on some of the quantum-mechanical results and the validity of the approximation itself can be tested. The result for the susceptibility can be found from that quoted for the quantum-mechanical method by dropping all but the highest power of S in each a_n. The values of kT_c/J are listed in Table I under the "K-O Method, Class., Roots" and ". . . Ratios."

The classical approximation is more useful when applied to the B-P-W method. Using the cluster Hamiltonian we have

$$Z_{cl}=\int \exp(-\mathcal{H}_{cl}/kT)d\Gamma_{cl}=\int d\Omega_0 \exp(\lambda_0\cdot\mathbf{S}_0)P^n, \tag{21}$$

where

$$P=\int d\Omega_i \exp(2\rho\mathbf{S}_0\cdot\mathbf{S}_i+\lambda_1\cdot\mathbf{S}_i). \tag{22}$$

To evaluate P, consider $\mathbf{S}_0$ to be the polar axis and integrate over the directions of $\mathbf{S}_i$. If ϑ and ψ are the angles made by $\mathbf{S}_i$ with $\mathbf{S}_0$ and $\mathbf{H}_1$, respectively, we obtain

$$P=\int d\Omega_i \exp(2\rho S^2\cos\vartheta+\lambda_1 S\cos\psi). \tag{23}$$

Assuming $\mathbf{H}_0$ and $\mathbf{H}_1$ to be parallel and letting η be the angle between $\mathbf{S}_0$ and $\mathbf{H}_0$, we have

$$\cos\psi=\cos\eta\cos\vartheta+\sin\eta\sin\vartheta\cos(\phi_1-\phi_0),$$

where ϕ_1 and ϕ_0 are the azimuthal angles of $\mathbf{S}_1$ and $\mathbf{H}_0$ in the plane perpendicular to $\mathbf{S}_0$.

For the purposes of locating the Curie temperature, it is sufficient to evaluate Z_{cl} to terms in λ_1^2 so that we need only

$$P=\int d\Omega_i \exp(2\rho S^2\cos\vartheta)\{1+\lambda_1 S\cos\psi+\tfrac{1}{2}(\lambda_1 S)^2\cos^2\psi\}. \tag{24}$$

The integrations are elementary and give

$$P=4\pi\{\alpha+\lambda_1 S\alpha'\cos\eta+\tfrac{1}{2}(\lambda_1 S)^2[\alpha''\cos^2\eta+\tfrac{1}{2}(\alpha-\alpha'')\sin^2\eta]\} \tag{25}$$

from which we have to first order in λ_0 and second order in λ_1,

$$Z_{cl}=\text{const}\alpha^n\{1+\tfrac{1}{2}\lambda_0\lambda_1 n S^2\mathcal{L}+\tfrac{1}{2}\lambda_1^2 n S^2[1+(n-1)\mathcal{L}^2]\}$$

where $\alpha=\sinh 2\rho S^2/2\rho S^2$ and

$$\mathcal{L}=\alpha'/\alpha=\coth 2\rho S^2-1/2\rho S^2$$
$$=\text{Langevin function}.$$

The zero-field magnetizations of a central and a n.n. atom are

$$m_0=\tfrac{1}{2}g\beta S^2\lambda_1 n\mathcal{L}, \tag{26}$$

$$m_1=\tfrac{1}{2}g\beta S^2\lambda_1[1+(n-1)\mathcal{L}^2]. \tag{27}$$

The consistency conditions require

$$C_0{}^F \qquad (n-1)\mathcal{L}^2-n\mathcal{L}+1=0, \tag{28}$$

$$C_0{}^{AF} \qquad (n-1)\mathcal{L}^2+n\mathcal{L}+1=0, \tag{29}$$

which are the same if we note that $\mathcal{L}(X)$ is an odd function. The roots of the quadratics are

$$\mathcal{L}(2|p_c|S^2)=1, \tag{30}$$

and

$$\mathcal{L}(2|p_c|S^2)=1/(n-1), \tag{31}$$

the first of which implies $p_c=\infty$ or $T_c=0$ and corresponds to the "anti-Curie point" found by Anderson[14] in the quantum-mechanical case. The other root gives the following values:

n	4 (q.l.)	6 (s.c.)	8 (b.c.c.)
$2\|J\|S^2/kT_c$	1.08	0.615	0.434.

The values of kT_c/J, which are listed in Table I, under "B-P-W method, Class, Exact," are seen to be, except for the q.l. lattice, reasonably close to the quantum-mechanical results for spin ≥ 1. The existence of a root for $n=4$ is interesting since the quantum-mechanical calculation does not predict a transition (for $S=\frac{1}{2}$ and 1) and the Ising model[4] (B-P-W method) does.

Before leaving the classical model we will mention one more approximation which enables us to find the cluster partition function more easily and more completely than above. As before, let $\mathbf{S}_1=\sum \mathbf{S}_i$. The element

[14] P. W. Anderson, Phys. Rev. **80**, 922 (1950).

of volume in phase space can then be written as

$$d\Gamma_{cl}=d\Omega_0 W(S_1)S_1^2 dS_1 d\Omega_1,$$

where $W(S_1)$ is the probability of forming a total spin S_1 which, for classical vectors, is the result of a random walk and can be calculated without great difficulty for small n, leading to the same partition function already found. For large n, however, the "asymptotic" limit of the random walk formula is[15] $W(S_1)=(q/\sqrt{\pi})^3 \times\exp(-q^2S_1^2)$, where $q=(3/2nS^2)^{\frac{1}{2}}$. It is now possible to evaluate Z_{cl} as a closed function of J/kT and H_1/kT. The result is

$$Z_{cl}=4\pi\exp(u^2+h^2)(\alpha+\lambda_0 S\alpha')$$

where

$$\alpha=\alpha(2uh)=\sinh 2uh/2uh, \quad u=pS/q, \quad h=\lambda_1/2q.$$

The Curie temperatures are given by the roots of $\mathfrak{L}(2uh)\mp h/(A-u)=0$, $(A=qnS=(3n/2)^{\frac{1}{2}})$ for small h, namely, $J/kT_c=(3/4S^2)\{1-(1-4/n)^{\frac{1}{2}}\}$, which is of the same form as Heisenberg's result—as it should be, since the random walk formula is Gaussian. This result is exactly the same as found by the second approximation in the classical K-O method or in the limit $S\to\infty$ of the quantum-mechanical K-O method [Eq. (12)]. Numerical values of kT_c/J are given in Table I under "B-P-W Method, Class, Asymptotic." The classical B-P-W method can be used to calculate other properties of the transition. Further results will be published later.

V. CONCLUSIONS

The mass of numerical data obtained allows certain conclusions to be drawn regarding the two methods employed. In brief, these conclusions are as follows:

The K-O method, when applied to ferromagnetic lattices, leads to sequences of approximations which are rapidly convergent, especially for large f and S. The limit of convergence can usually be located within a few percent and is apparently unique. Moreover, the convergence limits for the b.c.c. and s.c. lattices agree satisfactorily with the results of the B-P-W method. Since the B-P-W and K-O methods begin with the same Hamiltonian and arrive at solutions by different, and somewhat complementary types of approximation, the validity of both approximations for three-dimensional ferromagnetic lattices is evident from the agreement of the results.

For the q.l. lattice, dissimilar results are found. The B-P-W method predicts no transition for this lattice (or for the h.l. lattice[4,5]) except in the limit of classical spin. (This is in agreement with the Bloch theory according to which only three-dimensional lattices are ferromagnetic.) However, the K-O method leads to sequences of approximations that behave normally for both the h.l. and q.l. lattices, implying that a transition temperature exists.

[15] S. Chandrasekhar, Revs. Modern Phys. **15**, 1 (1943).

For antiferromagnetics, the sequences of approximations obtained by the K-O method do not have a unique convergence limit and the limits found do not agree with the results of the B-P-W method. It is suggested that the antiferromagnetic susceptibility does not have a singularity at the Curie temperature and that the Opechowski method is therefore inapplicable.

We wish to thank the Numerical Analysis Laboratory at the University of Wisconsin for finding the solutions of Eq. (19). One of us (H.A.B.) is indebted to the Wisconsin Alumni Research Foundation and the Engineering Research Institute at the University of Michigan for financial assistance.

APPENDIX A

The derivation of the cluster partition function [Eq. (14)] will now be considered. Following Karplus and Schwinger,[13] we introduce the function

$$F(X)=e^{X(A+B)},$$

where, if

$$A=2p\mathbf{S}_0\cdot\mathbf{S}_1$$

and

$$B=\boldsymbol{\lambda}_0\cdot\mathbf{S}_0+\boldsymbol{\lambda}_1\cdot\mathbf{S}_1,$$
$$Z_{cl}=\mathrm{trace}F(1).$$

The defining equation for F leads to the differential equation $(d/dX)F(X)=(A+B)F(X)$ and the boundary condition $F(0)=1$. A function $G(X)$ defined by

$$F(X)=e^{XA}G(X),$$

then satisfies the differential equation

$$(d/dX)G(X)=e^{-XA}Be^{XA}G(X),$$

with the boundary condition, $G(0)=1$. This is equivalent to the integral equation,

$$G(X)=1+\int_0^X e^{-X'A}Be^{X'A}G(X')dX',$$

which is our starting point. Choosing as a zeroth approximation, $G(X')=1$, a double iteration leads to

$$G(X)=1+\int_0^X e^{-X'A}Be^{X'A}dX'+\int_0^X e^{-X'A}Be^{X'A}\int_0^{X'} e^{-X''A}Be^{X''A}dX''dX',$$

and

$$F(1)=e^A+\int_0^1 e^{(1-X')A}Be^{X'A}dX'+\int_0^1 e^{(1-X')A}Be^{X'A}\int_0^{X'} e^{-X''A}Be^{X''A}dX''dX'.$$

Using the S_0S_1SM representation (in more usual notation, the $LSJM$ scheme) with $\mathbf{S}=\mathbf{S}_0+\mathbf{S}_1$, we have

$$\text{trace}F(1)=\sum_{S_1} w(S_1)\sum_S\sum_M F(1),$$

which, for convenience, we write as

$$\text{trace}F(1)=\sum_{S_1} w(S_1)\ \text{trace}F(1).$$

We have then

$$\text{trace}F(1)=f_0+f_1+f_2,$$

where

$$f_0=\text{trace}e^A=\sum_S\sum_M(SM|e^A|SM)$$
$$=\sum_S(2S+1)e^\alpha,$$

where $\alpha=$ eigenvalues of

$$A=p[S(S+1)-S_0(S_0+1)-S_1(S_1+1)]$$

and

$$f_1=\text{trace}\int_0^1 e^{(1-X')A}Be^{X'A}dX'$$

$$=\int_0^1 dX'\sum_S\sum_M(SM|e^{(1-X')A}Be^{X'A}|SM)$$

$=0$ since B is an odd function of M,

and

$$f_2=\text{trace}\int_0^1\int_0^{X'} dX'dX''e^{(1-X')A}Be^{(X'-X'')A}Be^{X''A}$$

$$=\text{trace}\int_0^1\int_0^{X'} dX'dX''Be^{(1+X''-X')A}Be^{(X'-X'')A}.$$

Transforming variables by means of

$$X'-X''=u,\quad X'+X''=v,$$

and performing the integration over v, we arrive at

$$f_2=\int_0^1 du(1-u)\ \text{trace}(Be^{(1-u)A}Be^{uA}).$$

Replacing $1-u$ by u leads to the integral

$$f_2=\int_0^1 udu\ \text{trace}(Be^{(1-u)A}Be^{uA}).$$

Taking one-half of the sum of these two expressions gives us

$$f_2=\tfrac{1}{2}\int_0^1 du\ \text{trace}(Be^{(1-u)A}Be^{uA}).$$

In this representation, A is diagonal and B is the z-component of a vector which, by writing $\mathbf{T}=\mathbf{S}_0-\mathbf{S}_1$, takes the form

$$\mathbf{B}=\tfrac{1}{2}(\lambda_0+\lambda_1)\mathbf{S}+\tfrac{1}{2}(\lambda_0-\lambda_1)\mathbf{T}.$$

Then, since $\text{trace}P_sQ_s=\frac{1}{3}\ \text{trace}\mathbf{P}\cdot\mathbf{Q}$, we can write f_2 in the form:

$$f_2=(1/24)[(\lambda_0+\lambda_1)^2g_0+(\lambda_0^2-\lambda_1^2)g_1+(\lambda_0-\lambda_1)^2g_2],$$

where

$$g_0=\int_0^1 du\ \text{trace}(\mathbf{S}e^{(1-u)A}\cdot\mathbf{S}e^{uA}),$$

$$g_1=\int_0^1 du\ \text{trace}(\mathbf{S}e^{(1-u)A}\cdot\mathbf{T}+\mathbf{T}e^{(1-u)A}\cdot\mathbf{S}e^{uA}),$$

and

$$g_2=\int_0^1 du\ \text{trace}(\mathbf{T}e^{(1-u)A}\cdot\mathbf{T}e^{uA}).$$

Since S and A commute, we have

$$g_0=\int_0^1 du\ \text{trace}\mathbf{S}^2e^{uA}$$
$$=\sum_S S(S+1)(2S+1)e^\alpha.$$

For g_1, the identity, $\mathbf{S}\cdot\mathbf{T}=\mathbf{S}_0^2-\mathbf{S}_1^2$, leads to

$$g_1=\int_0^1 du\ \text{trace}2\mathbf{S}\cdot\mathbf{T}e^A$$
$$=2\sum_S(2S+1)[S_0(S_0+1)-S_1(S_1+1)]e^\alpha.$$

To obtain g_2, we make use of several sum rules for angular momentum matrices.[16] Since A is diagonal,

$$\text{trace}\mathbf{T}e^{(1-u)A}\cdot\mathbf{T}e^{uA}$$
$$=\sum_{SM}\sum_{S'M'}(SM|\mathbf{T}|S'M')e^{(1-u)\alpha'}\cdot(S'M'|\mathbf{T}|SM)e^{u\alpha}.$$

Using equation 13^31^1, 10^32b^1, and 10^32a of reference 21, this becomes

$$\text{trace}\mathbf{T}e^{(1-u)A}\cdot\mathbf{T}e^{uA}$$
$$=\sum_S\{(P(S)Q(S-1)/S)e^{p[S(S-1)-b]+2up\beta}$$
$$+((2S+1)a^2/S(S+1))e^{p[S(S+1)-b]}$$
$$+(P(S+1)Q(S)/(S+1))e^{p[(S+1)(S+2)-b]+2up(S+1)},$$

where

$$b=S_0(S_0+1)+S_1(S_1+1),\quad a=S_0(S_0+1)-S_1(S_1+1),$$

and

$$P(S)=(S-S_0+S_1)(S+S_0+S_1+1),$$
$$Q(S)=(S+S_0-S_1)(-S+S_0+S_1+1).$$

After performing this integration and combining terms we have

$$g_2=\sum_S e^\alpha\{(2S+1)a^2/S(S+1)+P(S)Q(S-1)/pS^2$$
$$-P(S+1)Q(S)/p(S+1)^2\}.$$

[16] E. U. Condon and G. Shortley, *Theory of Atomic Spectra* (Cambridge University Press, London, 1935), Chap. III.

Finally, adding g_0, g_1, and g_2 gives, after some rearrangement,

$$f_2=(1/24)\sum_s e^{\alpha}\{(2S+1)/S(S+1)[\lambda_0(S(S+1)+a)+\lambda_1(S(S+1)-a)]^2+(\lambda_0-\lambda_1)^2[P(S)Q(S-1)/pS^2-P(S+1)Q(S)/p(S+1)^2]\}.$$

By taking out the term $S=0(S_0=S_1)$ from the sum (if it occurs), and using the identities

$$P(S)Q(S-1)/S^2-P(S+1)Q(S)/(S+1)^2=(2S+1)\sigma_0\sigma_1/S(S+1) \quad \text{for } S\neq 0$$
$$=4S_0(S_0+1), \quad \text{for } S=0$$

and

$$a=S(S+1)-b=\sigma_0-2S_0(S_0+1),$$

where σ_0 and σ_1, are as defined in Eq. (14), we can write the cluster partition function as given there if we multiply through by the constant $e^{2p\beta_0(S_0+1)}$.

This result can also be derived by standard perturbation theory methods, considering B as the perturbation. The eigenvalues of $\mathcal{H}_{cl}$ are then

$$E_{cl}=\alpha+(\alpha|B|\alpha)+\sum_{\alpha'}{}'\frac{|(\alpha|B|\alpha')|^2}{\alpha-\alpha'}$$

to second order in B,

$$Z_{cl}=\sum s_1 w(S_1)\sum_\alpha e^{-E_{cl}/kT}$$
$$=\sum s_1 w(S_1)\sum_\alpha e^{\alpha}\left\{1+(\alpha|B|\alpha)+\tfrac{1}{2}(\alpha|B|\alpha)^2+\sum_{\alpha'}{}'\frac{|(\alpha|B|\alpha')|^2}{\alpha-\alpha'}\right\},$$

with $\sum_\alpha=\sum_S\sum_M$. Again employing the properties of angular momentum matrices we arrive, after a similar amount of manipulation, at the result found above.

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Nuclear Magnetic Resonance in Semiconductors. I. Exchange Broadening in InSb and GaSb

R. G. Shulman, J. M. Mays, and D. W. McCall
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received May 17, 1955)

Nuclear magnetic resonances have been observed for the more abundant nuclear species in the semiconductors InSb and GaSb. Broad lines have been observed and explained by a nuclear spin exchange mechanism where the interaction between the nuclei is of the form $E_{ij}=A_{ij}\mathbf{I}_i\cdot\mathbf{I}_j$. The exchange coupling involves the hyperfine interaction between nuclear spins and electron spins. Therefore it is very sensitive to the electron energy states in both valence and conduction bands. It is particularly sensitive to the energy surfaces far from the Fermi level and provides information about the band structure of these normally inaccessible regions.

1. INTRODUCTION

NUCLEAR magnetic resonance has proved to be a valuable means for studying metals and insulators.[1] Resonance width, shape, multiplet structure, intensity, frequency and relaxation times all depend upon the environment or lattice in which the nucleus is situated. Study of the resonances therefore yields information about the nuclear environment. Investigation of nuclear magnetic resonances in semiconductors is being undertaken with the double purpose of studying the materials themselves and of using the delicate control that can be exercised over various properties of semiconductors to contribute to the theory of nuclear magnetic resonance. Intensive research in the past several years has produced materials of extremely high purity with consequent advances in the understanding of their properties. Impurities can be introduced in known concentration and the resultant changes in electronic structure have been investigated by other means in previous studies. In this way one can control and vary the concentration of conduction electrons and paramagnetic impurities, a situation which makes nuclear resonance investigations exceedingly attractive. The electronic structures, intermediate in many respects between those of metals and insulators, should allow a variety of interactions between the nuclei and their lattice.

The nuclear magnetic resonances of dielectric substances have been extensively investigated and the lines observed are usually broadened by internuclear dipole-dipole interactions. Van Vleck[2] has shown that dipole-dipole broadening is a fundamental property of a crystal lattice and the effects are observed in metals as well as insulators. From Van Vleck's theory one can obtain lattice parameters in favorable instances. Pound[3] and Bloembergen[4] have shown quadrupole interactions

[1] H. S. Gutowsky, Ann. Rev. Phys. Chem. **5**, 333 (1953).

[2] J. H. Van Vleck, Phys. Rev. **74**, 1168 (1948).
[3] R. V. Pound, Phys. Rev. **79**, 685 (1950).
[4] N. Bloembergen, *Nuclear Magnetic Relaxation* (Martinus Nijhoff, The Hague, 1948).

to be a determining factor in the shape and width of nuclear magnetic resonances in some insulators. Multiplet structure in liquids has provided information on molecular configuration and chemical bonding.[6-7] This type of nuclear exchange interaction, first discussed by Ramsey and Purcell in covalent bonds, should also be effective in solids when the nuclei have large atomic numbers.

In metals the nuclei have been shown to interact with the conduction electrons, the neighboring nuclei, and to couple to the crystalline field gradients via nuclear quadrupole moments. Interactions with conduction electrons have been shown by Knight[8] to shift the nuclear resonance frequency from that observed in nonmetals. Korringa[9] has discussed spin-lattice relaxation in terms of this interaction and has related the frequency shift to relaxation times. Experiments by Gutowsky and McGarvey[10-12] have confirmed the conduction electron relaxation mechanism for several metals. Bloembergen and Rowland,[13] in a summary of nuclear resonance in metals, have shown the importance of electric quadrupole interactions as well as illustrating in detail some of the effects of conduction electrons.

Ruderman and Kittel[14] have recently proposed that internuclear exchange coupling[5] is important in metallic silver and they have derived the general theory for metals. This theory is applicable to other metals and the effect will be important when the atomic number is high. Bloembergen and Rowland[15] have employed this theory in their analysis of nuclear resonance experiments on metallic thallium. They have further derived the corresponding theory for insulators and applied this analysis to thallic oxide.

In the present paper we present the results of a nuclear magnetic resonance investigation of the semiconductors InSb and GaSb. Among other things it will be shown that the internuclear exchange interaction is dominant in determining the resonance widths and shapes in these compounds. Anderson[16] has extended the work of Ruderman and Kittel[14] and Bloembergen and Rowland[15] to the case of semiconductors. Fundamental information regarding the band structure of semiconductors is obtainable from the nuclear resonance measurements.

II. APPARATUS

The resonances were observed with a Pound-Knight-Watkins spectrometer.[17,18] The rf unit, 280 cycle modulation generator, and lock-in amplifier were built from Watkins' diagrams. The rf unit is powered by a Lambda Model 25 regulated power supply and a dc filament supply. Frequency measurements are made with a BC-221 frequency meter and the oscillator is monitored with a National HRO-60 communications receiver. Resonances are displayed on a Leeds and Northrup Speedomax recorder with a -5 to 0 to $+5$ millivolt range and a one-second pen. The lock-in amplifier is designed to allow observation with 2, 8, or 18 sec time constants. The 280 cycle sine-wave modulation is introduced by two coils taped to the magnet pole-faces and is variable from 0–5 gauss peak-to-peak. The sample coil is mounted in a cryostat, designed by Dr. W. P. Slichter, similar to that described by Gutowsky *et al.*[19] Measurements at liquid nitrogen temperature were made by actually submerging the sample coil assembly in the nitrogen. The compensation coil, used to balance out 280-cycle pickup, consists of a double layer, three inches long, of No. 30 wire wound around the outside of the cryostat dewar.

A permanent magnet with a field of approximately 6520 gauss in a 2-inch gap 6 inches in diameter was used. The magnet was designed and constructed by the Indiana Steel Products Company, Valparaiso, Indiana, under the direction of Mr. H. L. Johnson. The design is similar to that of the magnet described by Gutowsky *et al.*[19] Proton resonances observed in water show that homogeneity at the center of the magnet gap is at least as good as 0.5 gauss over the 1-cc samples used. The field can be swept over a range of about 120 gauss by means of coils of 2000 turns of No. 20 wire around each Alnico pole. Current varying linearly in time is provided by an electronic regulator[19] and measured with a Sensitive Research Instrument Corporation Model C dc milliammeter. This meter, with ten overlapping ranges from 1.5 to 1500 ma full scale, is very useful for both high resolution and broad line measurements.

III. EXPERIMENTAL RESULTS

A. Gallium Antimonide

In gallium antimonide resonances of Ga^{69}, Ga^{71}, Sb^{121}, and Sb^{123} were observed. Table I lists pertinent data for the nuclei under consideration. These data are taken from the table published by Varian Associates. Two different samples of the highest purity obtainable

[5] N. F. Ramsey and E. M. Purcell, Phys. Rev. **85**, 143 (1952).
[6] Gutowsky, McCall, and Slichter, J. Chem. Phys. **21**, 279 (1953).
[7] N. F. Ramsey, Phys. Rev. **91**, 303 (1953).
[8] W. O. Knight, Phys. Rev. **76**, 1259 (1949).
[9] J. Korringa, Physica **16**, 601 (1950).
[10] H. S. Gutowsky and B. R. McGarvey, J. Chem. Phys. **20**, 1472 (1952).
[11] B. R. McGarvey and H. S. Gutowsky, J. Chem. Phys. **21**, 2114 (1953).
[12] H. S. Gutowsky, Phys. Rev. **83**, 1073 (1951).
[13] N. Bloembergen and T. J. Rowland, Acta Metallurgica **1**, 731 (1953).
[14] M. Ruderman and C. Kittel, Phys. Rev. **96**, 99 (1954).
[15] N. Bloembergen and T. J. Rowland, Office of Naval Research Technical Report 205, October 5, 1954 (unpublished). We are indebted to Dr. Bloembergen for kindly providing us with a copy of this report.
[16] P. W. Anderson (private communication).
[17] R. V. Pound and W. D. Knight, Rev. Sci. Instr. **21**, 219 (1950).
[18] George D. Watkins, thesis, Harvard University, 1952 (unpublished).
[19] Gutowsky, Meyer, and McClure, Rev. Sci. Instr. **24**, 644 (1953).

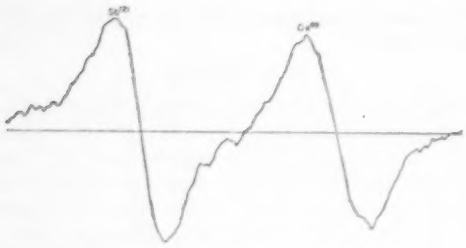


FIG. 1. Typical recorder traces of the derivative of the Ga^{69} and Sb^{121} resonances.

were used. These samples, as well as those of indium antimonide, were put at our disposal by Dr. H. J. Hrostowski, of these Laboratories, who also kindly supplied the analysis of donor and acceptor concentrations. Both GaSb samples were p-type with a hole concentration of about $10^{17}/cm^3$ at room temperature and at liquid nitrogen temperature. Because of the relatively high resistivity ($\sim$0.06 ohm cm), the skin depth at 8 Mc/sec was about 0.02 cm. Powder samples could easily be prepared and relatively large single crystals could be used with sufficient penetration of the radio-frequency energy. The sample used for most of the measurements was a powder prepared from a zone refined polycrystalline bar. In order to determine the orientation dependence of the resonance a single crystal with the (100) direction along the axis of the coil was used. Since results, except for signal to noise ratios, were the same for both the single crystal and the powder samples and were furthermore not temperature dependent between 77°K and 300°K, we report only the results on the powder at 77°K.

Typical recorder traces of the derivative of the Ga^{69} and Sb^{121} resonances are shown in Fig. 1. The lines were obtained by sweeping the magnetic field and the peak-to-peak widths are five to six gauss.

Although no very accurate measurements of the central frequency of these broad lines were possible, no shifts of the central frequencies from the values listed in the Varian Associates table were detected within an experimental accuracy of 2 kc/sec in 6–8 Mc/sec. Line shape measurements are presented in Table II. Column 2 lists the separation, in gauss, of the points of maximum deflection of the derivative curves plotted. The values quoted are the average of several determinations and limits of experimental error are presented.

The experimental absorption second moments of the lines are listed in column 3. These are average values from several traces and are calculated from the formula,[20]

$$\Delta H_2^2=(1/3)\int_{-\infty}^{\infty}(H-H_0)^3\left(\frac{dI}{dH}\right)dH\Big/\int_{-\infty}^{\infty}\times(H-H_0)\left(\frac{dI}{dH}\right)dH.$$

[20] G. E. Pake and E. M. Purcell, Phys. Rev. **74**, 1184 (1948).

The precision of the second-moment determinations is not as good as that obtained for the line widths owing to the relative importance of the tails of the trace. Also errors in the second moments are introduced by base line drift and the difficulty in placing the line center. Close agreement between the average values for the two halves of the lines confirmed the visual evaluation that the lines are symmetrical. Even though the $Ga^{69}-Sb^{121}$ pair almost overlaps, the individual lines look fairly symmetrical and the second moments for the upper and lower frequency halves agree. However additional difficulty in selecting a base line was encountered and the data are not highly reliable.

In column 4, theoretical second moments calculated from Van Vleck's[2] equation for dipole-dipole broadening are given.

$$\Delta H_2^2=\left(\frac{3}{5}\right)I(I+1)\gamma^2\hbar^2\sum_j r_{0j}^{-6}+\left(\frac{4}{15}\right)\sum_f I_f(I_f+1)\gamma_f^2\hbar^2 r_{0f}^{-6}. \quad (1)$$

Lattice constants were taken from the *Handbook of Chemistry and Physics*.[21]

TABLE I. Pertinent nuclear data.

Nucleus	Frequency (mc/sec)	Field (gauss)	Spin	Quadrupole moment (10^{-24} cm^2)	% Abundance
Ga^{69}	6.67	6520	3/2	0.2318	60.2
Ga^{71}	8.47	6520	3/2	0.1461	39.8
In^{113}	6.06	6520	9/2	1.144	4.16
In^{115}	6.08	6520	9/2	1.161	95.84
Sb^{121}	6.64	6520	5/2	−1.3	57.25
Sb^{123}	3.60	6520	7/2	−1.7	42.75

The fifth column presents the experimental integrated intensities in the same arbitrary units used for the InSb intensities.[22] These are compared with the theoretical values in the last column. The permanent magnetic field was homogeneous to within $\pm\frac{1}{2}$ gauss and the modulation was kept small compared with the line width in order to eliminate modulation broadening.

[21] *Handbook of Chemistry and Physics* (Chemical Rubber Publishing Company, Cleveland, 1948), thirtieth edition.

[22] In these units D, the recorder deflection, is measured in volts by the calibrator, H is the magnetic field and γ the nuclear gyromagnetic ratio. In order to express the integrated intensity in terms of recorded quantities we integrate by parts.

$$\int_0^\infty I d\nu = I\nu\Big|_0^\infty - \int_0^\infty \nu dI.$$

For a reasonable line shape $I\nu|_0^\infty = 0$ so

$$\int_0^\infty I d\nu = -\int_0^\infty \nu dI = -\gamma\int_0^\infty H dI = -\gamma\int_0^\infty H\left(\frac{dI}{dH}\right)dH.$$

Now $D=\left(\frac{dI}{dH}\right)\Delta H_m$ where ΔH_m is the modulation width. Therefore

$$\int_0^\infty I d\nu \propto \frac{\gamma}{\Delta H_m}\int_0^\infty H D dH.$$

Saturation effects were noticed in Ga^{71} and Ga^{69}, but not in Sb^{121}. Observations reported were all taken at the lowest possible oscillator power level. As the power level was increased saturation appeared in the Ga^{71} resonance before it was seen in the Ga^{69} while within the power limits available the Sb^{121} did not change.

B. Indium Antimonide

In indium antimonide resonances of Sb^{121} and In^{115} were observed. The In^{113} is only 4.16% abundant and was too weak to be detected. The Sb^{123} resonance was not seen, presumably because it was too weak and too broad. The In^{115} resonance on the other hand, had the largest signal-to-noise ratio of any of the lines observed in either of the two compounds.

The InSb samples used were purer than the GaSb samples. In GaSb, the density of impurities acting as acceptors was $\sim 10^{17}/cm^3$ and the total impurity concentration was at least this high. InSb had net acceptor impurity concentrations of $\sim 10^{15}/cm^3$ in the two samples studied most extensively. These samples were intrinsic conductors at room temperature containing

TABLE II. Resonance, widths, second moments, and intensities.

Nucleus		δH (gauss)	Experimental ΔH_2^2 (gauss^2)	ΔH_2^2 (gauss^2) Calc. Eq. (1)	Theoretical intensity[a]	Experimental intensity
GaSb	Ga^{69}	5.1 ±0.14	6.5	1.20	0.020	0.024
	Ga^{71}	5.58±0.1	6.2±0.3	1.20	0.028	0.061
	Sb^{121}	4.7 ±0.1	6.7	1.05	0.045	0.024
	Sb^{123}	5.1 ±0.2	8.4±0.9	0.93	0.020	
InSb	In^{115}	9.0 ±0.2	24±1	1.65	0.167	0.167
	Sb^{121}	17.5 ±0.6	65±4	2.52	0.045	0.034

[a] Taken from the Varian N.M.R. Table, constant field. The experimental values are given in arbitrary units the In^{115} value being adjusted to the theoretical value.

$\sim 3\times 10^{16}/cm^3$ electrons and holes. At 77°K, the other operating temperature, they are extrinsic, p-type semiconductors with a mobile hole concentration of $\sim 3\times 10^{15}/cm^3$. Despite the change in carrier concentration by an order of magnitude between these two temperatures, no line shape changes were observed.

No differences were observed between the line shapes in a single crystal sample and a polycrystalline powder. The widths between extrema for the In^{115} resonance, $\delta H(In^{115})$, in the single crystal, were measured as a function of angle of orientation. The crystal was rotated in increments of 30° about an axis along the ⟨100⟩ direction which was also the coil axis. It was seen that $\delta H(In^{115})$ was equal to the powder value and independent of orientation. Both the In^{115} and Sb^{121} lines were Gaussian in shape. The In^{115} line for the powder sample at 77°K is shown in Fig. 2. The Gaussian shape function $C\exp[-K(H-H_0)^2]$ is plotted as crosses and fits the recorder trace quite well. The circles are a plot of the Lorentzian shape function $C/[1+(H-H_0)^2]$ for comparison. Both shape functions were made to pass through the center point and the maximum. Comparison is made for only one-half the line because of slight distortion of the other half by the In^{113} resonance.

Other experimental results are recorded in Table II. Owing to the great strength of the In^{115} its results are the most reproducible. The Sb^{121} resonance, on the other hand, was the weakest observed and its properties are not very accurately determined, particularly the second moment. No saturation effects were seen.

In addition, a single crystal of InSb was measured and then plastically deformed 20% by E. S. Greiner, of these Laboratories. The In^{115} resonance, observed before and after compression, showed a reduction in intensity by a factor of ~4.7. Although the In^{115} resonance was weak, there was no clear indication that it had been broadened or distorted. The 4.7 fold reduction represents the reduction of the peak derivative signal.

IV. DIPOLE-DIPOLE, QUADRUPOLE, AND T_1 EFFECTS

The most striking experimental results are the widths of the absorption lines. These line widths will be explained below by a nuclear spin exchange mechanism. Before this, however, it is necessary to examine several more traditional mechanisms which might determine the line widths and shapes. These are dipolar broadening and first and second order quadrupole interactions.

We see from Table II that the measured second moments are from five to 35 times larger than those calculated from Eq. (1). Since this relation is an exact expression for the contributions of dipolar broadening to the second moment we must look for additional mechanisms to explain the observations. Further evidence that dipolar broadening is not important in indium and gallium antimonides is furnished by the experiments on the orientation dependence of the line widths in single crystals. No change in width within experimental error was observed in the In^{115} resonance in InSb on rotation of the sample in 30° increments while the theoretical second moment calculated from Van Vleck's[2] Eq. (11) changes by a factor of 3.5.

All the nuclei observed have large electric quadrupole moments. Both InSb and GaSb have wurtzite struc-

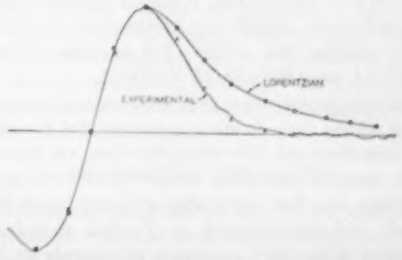


FIG. 2. Typical recorder trace of the derivative of the In^{115} resonance. Crosses are plotted from the Gaussian shape function and circles from the Lorentzian shape function.

tures where each atom is surrounded by four tetrahedral bonds to atoms of the other type. In a perfect crystal of this symmetry, there should be no electric field gradient at the nuclei. However, it is well known that actual crystals are not perfect and it has been demonstrated that even cubic crystals exhibit quadrupole interactions. Watkins and Pound[23] have shown that the integrated absorption intensities in KBr and KI are smaller than predicted by theory. They attribute this reduction to quadrupole splitting where random electric field gradients were established by lattice imperfections.

Pound[3] and Bersohn[24] have carried through perturbation theory treatments of quadrupole interactions in crystals. The resultant equation expresses the energy of the nuclear state as a power series in $(e^2Qq/\gamma\hbar H)$ where e is the proton charge, Q the nuclear quadrupole moment and q the electric field gradient at the nucleus. This fraction is small compared with unity in cases of interest in nuclear magnetic resonance and only the terms proportional to the first and second powers of $(e^2Qq/\gamma\hbar H)$, i.e., the first- and second-order terms, need be considered here. If the observed broadening is caused by the first-order terms then $(e^2Qq/\gamma\hbar H)$ is of the order of 10^{-3} and the second-order terms are negligible. The $m_I=\frac{1}{2}\leftrightarrow-\frac{1}{2}$ transitions are unaffected by first order quadrupole effects and therefore if the line shapes were the result of first-order broadening a sharp central spike would remain. The Gaussian shapes observed would require an extremely unlikely distribution of field gradients q. If the broadening were caused by the second-order quadrupole terms then $(e^2Qq/\gamma\hbar H)^2$ is of the order of 10^{-3} and the first-order splitting would be so great that only the $m_I=\frac{1}{2}\leftrightarrow-\frac{1}{2}$ transitions would be observable. As second-order quadrupole broadening results in unsymmetrical lines it is inconsistent with the present observations. Furthermore, the reduction in intensity of the In^{115} resonance in InSb after plastic deformation is proof that most of the components have not been split out in the undeformed crystal. We observe an intensity decrease by a factor of 4.7 in this case whereas a decrease by a factor of 6.6 would result if no quadrupole splitting was originally present and only the $m_I=\frac{1}{2}\leftrightarrow-\frac{1}{2}$ components remained after deformation. A further argument against the quadrupole broadening mechanism is provided by the relative line widths. For example, $Q(Ga^{69})/Q(Ga^{71})>1$, yet $\delta H(Ga^{69})/\delta H(Ga^{71})<1$.

Although quadrupole interactions are not responsible for line widths they do seem to account for the relaxation times observed. The relaxation time was measured by the onset of saturation in Ga^{69} to be $\sim$0.1 second. Saturation was first noticeable at lower power levels in Ga^{71} and corresponded to $T_1\sim0.2$ seconds. No tendencies of the Sb^{121} resonance to saturate could be observed indicating a shorter relaxation time. These values are roughly proportional to the quadrupole moments.

Finally it is probable that the deviation of relative intensities from the predicted values (see Table II) is due to quadrupole splitting, but not enough experimental evidence is available to corroborate this contention at this time.

V. PHYSICAL MODEL AND THEORY

Nuclear spin exchange via electron interactions was first described by Ramsey and Purcell[5] and applied to covalent bonds. Their theory has recently been modified and applied to metals by Ruderman and Kittel.[14] Regardless of the type of substance the basic mechanism is the hyperfine interaction of electron spin moments with nuclear spin moments. An electron in the vicinity of a nuclear spin moment μ_i will be polarized by that moment. The polarized electron then presents an effective field at another nucleus designated μ_j. By the same kind of hyperfine interaction μ_j feels the effect of the net polarization and in this fashion interacts with μ_i. It is apparent that for this effect to be large the electron must spend an appreciable time in the vicinity of both nuclei. Therefore we will be concerned with binding electrons only. The strength of this interaction will depend upon the magnitudes of μ_i and μ_j as well as upon other parameters which will be discussed in more detail below. It is possible to describe a polarization of electron spin states mathematically in terms of perturbation theory. The unperturbed, unpolarized electrons may be described by a wave function ψ_0 while the polarized case will correspond to mixing contributions from other states. The final function will represent the electron spin polarization by including wave functions in which the electron spins are aligned. Because of the large separation of electron energy states compared to thermal energy it is usually a good approximation to assume that all electronic states below a certain energy are filled while those above are empty. Furthermore, in the filled states it will be found that wave functions corresponding to both orientations of the electron spin are occupied by electrons. Therefore in perturbing the occupied states (where the electrons are found and hence the states which must be perturbed) we need only consider contributions from the unoccupied states. In solids, where one-electron wave functions are used, the exclusion principle does not influence the exchange either through the nuclear functions or the electronic functions but only through its determining influence upon the population of states. This follows from the fact that the Bloch functions for electrons in a solid are traveling waves and hence the interaction between μ_i and μ_j is essentially a double scattering process rather than an overlap of wave functions of electrons on each nucleus. This scattering is virtual rather than real because the linkage

[23] G. D. Watkins and R. V. Pound, Phys. Rev. **89**, 658 (1953).
[24] R. Bersohn, J. Chem. Phys. **20**, 1505 (1952).

involves a mixing of states which is a virtual excitation of electrons.

A quantitative expression for nuclear exchange interaction in semiconductors has been derived by Anderson[16] along the lines followed by Ruderman and Kittel[14] for metals. The highlights of Anderson's derivations are as follows. Assuming Bloch wave functions of the form $\varphi_k(\mathbf{r})=u_k(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$ and using Ramsey's perturbing Hamiltonian $\mathcal{H}_3$ the exchange interaction is

$$H_{ij}=-(\mathbf{S}\cdot\mathbf{I}_i)(\mathbf{S}\cdot\mathbf{I}_j)\Delta_i\Delta_j\sum_{\mathbf{k}=0}^{\infty}\sum_{\mathbf{k}'=0}^{\infty}\times\frac{\exp[-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{ij}]}{E_g+(h^2\mathbf{k}'^2/2m_e)+(h^2\mathbf{k}^2/2m_h)}, \quad (2)$$

$$\Delta_i\equiv\int u_{k'i}\left[\frac{8\pi}{3}\mu_i\mu_0\delta(\mathbf{r}_i)\right]u_{ki}d\tau,$$

where i and j refer to nuclei i and j, S is electron spin, I is nuclear spin, Δ_i is the matrix element of hyperfine interaction for the s electrons with atomic wave functions $u_k(\mathbf{r})$, and E_g is the energy gap of the semiconductor. This expression can be integrated if spherical energy surfaces are assumed and the result is $\Delta E_{ij}=A_{ij}\mathbf{I}_i\cdot\mathbf{I}_j$, where

$$A_{ij}=\frac{3.36\times10^{-7}\Omega^2m^*\xi_i\xi_j\psi_i^2(0)\psi_j^2(0)\mu_i\mu_j}{I_iI_jR_{ij}^4}\ \text{sec}^{-1} \quad (3)$$

is the first and most important term in the solution. Here Ω=atomic volume, $m^*=4(m_e^*)^{\frac{1}{2}}(m_h^*)^{\frac{1}{2}}/(m_e^*+m_h^*)^2$, $\psi_i^2(0)$=the probability of finding the outer s electrons of atom i at its nucleus, and

$$\xi_i=\frac{[\psi_i^*(0)_{\text{hole}}\psi_i(0)_{\text{electron}}]_{\text{solid}}}{[\psi_i^2(0)]_{\text{atom}}},$$

$$R_{ij}=|\mathbf{r}_i-\mathbf{r}_j|.$$

It is important to remember that this perturbation applies only to s electrons. It is assumed that the hyperfine interactions of p, d, etc. electrons are negligible compared to that of s electrons. Furthermore, it is important to note that the nuclear exchange calculated here only results in line broadening when the interaction is between unlike nuclei.

One interesting point about the integration over the energy surfaces is that the solution does not include the energy gap E_g in the dominant term. It does enter into the higher order terms but these are negligible for near-neighbor interactions. The reason for this is that the high momentum states have the greatest statistical weight and compared to their energy separations E_g can be ignored. For R_{ij} large compared to near-neighbor separations the high momentum states lose their phase coherence and no longer add constructively, but for small values of R_{ij} they do and are the important terms.

The limiting accuracy of the calculation is the integration over momentum space in both the valence and conduction bands. This integration will be discussed below in conjunction with an attempt to determine the absolute value of A_{ij}.

VI. LINE SHAPES

The shapes of the resonance lines will be calculated assuming the interaction discussed in the last section. Van Vleck has presented expressions for the second moments and fourth moments to be expected from exchange broadening. However, his expression for the fourth moments is difficult to evaluate and we have determined the theoretical line shapes not by comparing second and fourth moments but by numerically considering only nearest-neighbor interactions. We consider three interacting nuclear species, i, j, and k (e.g., In^{115}, Sb^{121}, and Sb^{123} respectively in InSb) and assume the resonance of nucleus i is to be observed. By a slight extension of the arguments presented by Gutowsky, McCall, and Slichter[6] the resonance of i is seen to occur at

$$\omega_i=\gamma_i[H_i+(C_{ij}\gamma_jM_j/\hbar)+(C_{ik}\gamma_kM_k/\hbar)], \quad (4)$$

where M_j and M_k are quantum numbers[25] for the spin groups of nuclei j and k and $C_{ij}\gamma_i\gamma_j\equiv A_{ij}$. The components thus occur at an equivalent magnetic field given by the bracketed term in Eq. (4). The constants C_{ij} and C_{ik} depend on the electronic properties of the same elements so they are equal, $C_{ij}=C_{ik}=C$. If K is written for γ_k/γ_j Eq. (4) becomes

$$\omega_i=\gamma_i[H_i+(C\gamma_j/\hbar)(M_j+KM_k)]. \quad (5)$$

With this equation the line shape may be calculated by computing the positions and intensities of the components and taking a suitable envelope of the resulting spectrum. The intensities are obtained in arbitrary units, depending only upon the relative weights of the spin states corresponding to M_j and M_k. The broadening is obtained in terms of the constant C which may then be found by comparison of the calculated and experimental curves. Line shapes for all of the resonances observed in InSb and GaSb have been calculated in the foregoing manner and are all Gaussian, in agreement with the experimental results. This Gaussian shape is to be expected considering the large numbers of nuclear spin orientations allowed. In the In^{115} Sb resonance, for example, there are 475 random components from nearest neighbor interactions alone.

VII. RELATIVE LINE WIDTHS

Following the procedure of Sec. VI one may obtain the resonance widths as a function of the field parameter $(C_{ij}\gamma_j/\hbar)$ by directly plotting the line shape. Therefore

[25] Let there be n_j nuclei in spin group j. Then if $F_j=n_jI_j$ we have $M_j=F_j, F_j-1, \cdots, -F_j$.

the relative line width may be obtained directly as the constant C_{ij} in any one compound cancels and γ_i and γ_j are known. This procedure assumes that only nearest neighbor interactions are appreciable as was previously justified through the rapid attenuation of A_{ij} with internuclear separation R_{ij}. Line width ratios have been computed in this way for InSb and GaSb and the results agree well with the experimental values.

A much simpler procedure, which allows the inclusion of second nearest neighbors, makes use of Van Vleck's equation (28).[2] Again the absolute value of A_{ij} is not required to determine the relative line widths, but we must use the radial dependence of A_{ij} which was shown (in Sec. V) to be an inverse fourth power. It will be shown in the following calculations that the line width contributions from second nearest neighbors are negligible in all the resonances reported here except for Sb in GaSb.

Van Vleck's equation (28) states

$$h^2\langle\Delta\nu_i^2\rangle_{Av}=\tfrac{1}{3}\textstyle\sum_f[I_f(I_f+1)\sum_j A_{ij}^2(R_{ij})], \qquad (6)$$

where the subscripts f refer to different nuclear species and the dipole-dipole terms have been omitted. Note that the dependence of A_{ij} on R_{ij} must be included. No interaction between like nuclei appears.

The comparison of relative line widths calculated by Eq. (6) and the experimental values is presented in Table III. The first column lists not the ratio of experimental line widths as measured but the values after correcting for dipolar broadening. The procedure followed was based upon the observed Gaussian line shapes. Since the peak-to-peak absorption derivatives were measured more precisely than the second moments these widths were converted to second moments by the relation $(1/4)\delta H^2=\Delta H_{\frac{1}{2}}^2$. Values of the second moment predicted from dipolar broadening were then subtracted from these experimental second moments and the results converted back to peak-to-peak absorption widths. The second column lists the theoretical ratios considering first and second nearest neighbors while the third column presents the effects of first nearest neighbors only. It can be seen that the antimony second nearest neighbors in GaSb have an appreciable effect. The $Sb^{121}-Sb^{123}$ exchange in these substances, while decreased compared to the GaSb exchange because of the larger separation, is increased, first by the larger number of second nearest neighbors (twelve), and second by a larger value of $\psi^2(0)$ for Sb compared to Ga. The values of $\psi^2(0)$ used will be discussed in the next section.

TABLE III. Relative line widths.

Nuclear ratio		Experimental width ratio, corrected	Theoretical ratio 1st and 2nd neighbor	Theoretical ratio 1st neighbor only
GaSb	$\frac{Ga^{71}}{Ga^{69}}$	1.11	1.00	1.00
GaSb	$\frac{Sb^{123}}{Sb^{121}}$	1.12	1.12	1.00
GaSb	$\frac{Ga^{71}}{Sb^{123}}$	1.08	0.99	1.21
InSb	$\frac{Sb^{121}}{In^{115}}$	2.00	1.73	1.73

VIII. ABSOLUTE VALUE OF A_{ij}

The energy surfaces in these compounds have previously been investigated in limited regions. Optical and Hall measurements of InSb[26] and GaSb[27] have been interpreted to yield values of the effective mass for holes and electrons. These measurements were made on electrons and holes occupying the bottom of the conduction band and the top of the valence band respectively. They indicate spherical energy surfaces in these regions. Values of the effective masses are listed in Table IV. However, the important contributions to the nuclear exchange come from the high momentum states in both bands, regions about which the experiments mentioned above tell us nothing.

TABLE IV. Effective masses of carriers in InSb and GaSb near the forbidden gap.

	m_e^*	m_h^*
InSb	0.025[a]	0.15[a]
GaSb	0.2[b]	0.4[b]

[a] H. J. Hrostowski (unpublished); more recent results indicate slightly different values.
[b] H. Leifer and W. C. Dunlap, Phys. Rev. **95**, 51 (1954).

The band structure of these compounds can be considered as arising from perturbations of the atomic functions. Isolated gallium, indium or antimony atoms have eight eigenstates in their outer shells. When compounds are formed the bonding orbitals are sp^3. There are four degenerate sp^3 bonding orbitals, and four degenerate sp^3 nonbonding orbitals. The former are the basis for the valence bands in the solid and the latter go into the first conduction band. In the solid the bonding and nonbonding orbitals are each split into four bands by interatomic interactions. The extent to which these bands are separated, and the shapes of the bands, are not known although calculations similar to those made for silicon and germanium[28] presumably could supply a great deal of information about the band structure. The magnitude of the interaction constant represents an experimental result which must be consistent with any calculation of the energy band surfaces. To show that the experimental results can be

[26] H. J. Hrostowski (private communication).
[27] H. Leifer and W. C. Dunlap, Phys. Rev. **95**, 51 (1954).
[28] F. Herman, Phys. Rev. **93**, 1214 (1954).

obtained with reasonable assumptions, consider two unlikely extremes. First we consider the minimum splitting case where the small measured values of effective mass are assumed to prevail throughout the entire band and where only one nondegenerate band lies low enough in energy to contribute appreciably. This corresponds to the electrons having only one quarter s character and the high momentum states in both valence and conduction bands possessing very large energies which increase the denominator of Eq. (2). The results of this approximation are listed in the first column of Table V and are seen to be smaller than the experimental values of 9.0 gauss for InSb and 5.6 gauss for GaSb. In order to calculate an extreme upper limit for the magnitude of A_{ij} we assume that the bands are degenerate everywhere, i.e., all four bands contribute equally, the effective masses equal the free electron mass, and that each atom is surrounded by four electrons with contributions from an s state and three p states, the latter being neglected. These values are listed in the last column. It is seen that the exchange interaction is quite sensitive to the energy surfaces while the measured values are reasonably located between the two extremes.

C. Herring has mentioned to the authors that in these lattices the "empty lattice" approximation should represent a reasonable description of the carriers over a major portion of the energy bands. By assuming that each electron and hole has the free electron mass, one quarter s character, and values of $\xi_i=\xi_j=0.74$ for InSb and $\xi_i=\xi_j=0.88$ for GaSb, agreement of Eq. (3) with the experimental line widths is obtained, showing the reasonable nature of this assumption.

It is an interesting coincidence that Eq. (11) of reference 14 gives exactly the same value of A_{ij} for InSb as our Eq. (3).

It is necessary to know the values of $\psi^2(0)$ for s-electrons in atomic gallium, indium and antimony. For gallium and indium measurements of hyperfine splitting by the $4s$ and $5s$ electrons in the doubly ionized state are available.[29] By comparing these values with those observed for the $4s$ and $5s$ electrons in copper and silver,[29] compromise values of $\psi_{Ga}^2(0)=8\times10^{25}$ cm^{-3} and $\psi_{In}^2(0)=16\times10^{25}$ cm^{-3} were used. Since no direct measurements had been made on Sb^{+4} the Fermi-Segrè formula was used in conjunction with a comparison with $\psi_{In}^2(0)$ and the value 24×10^{25} cm^{-3} appeared and was used for $\psi_{Sb}^2(0)$.

The exchange interaction between two nonidentical nuclei has been calculated from second-order perturbation theory. In the calculation it has been assumed that the exchange interaction is much smaller than the differences in resonance frequencies or that $A_{ij}\ll[|\gamma_i|-|\gamma_j|]H_0$. For the $Ga^{69}-Sb^{121}$ pair in GaSb at 6520 gauss this condition is just satisfied since the experimental value of $A_{ij}=0.8$ gauss while the measured value of the separation is 18.0 gauss. By varying the magnetic field strength it should be possible to change the separation and see how the lines collapse into a single line at small values of H_0, as predicted by Anderson.[30] It is planned to do this in the future.

[29] Landolt-Börnstein, *Physikalisch-Chemische Tabellen* (Springer-Verlag, Berlin, 1952), Vol. 1, Part 5, pp. 19, 29.

TABLE V. Comparison of experimental and theoretical interaction constants.

	Minimum splitting case assuming	Experimental	Maximum splitting case assuming
	(1) m^* InSb $=0.03m$ m^* GaSb $=0.254m$		(1) m^* InSb $=m^*$ GaSb $=1$
	(2) Only one band contributes to each integration over k space of $\xi_i=\xi_j=\frac{1}{4}$		(2) Four bands contribute equally to each integration over k space or $\xi_i=\xi_j=1$
$\delta H_{\frac{1}{2}}$ $In^{115}Sb$ (gauss)	0.5	9.0	260
$\delta H_{\frac{1}{2}}$ $Ga^{71}Sb$ (gauss)	1.8	5.6	115

CONCLUSIONS

The nuclear magnetic resonances in InSb and GaSb have been observed and the following conclusions reached.

(1) Line widths and shapes are almost completely determined by a nuclear exchange mechanism involving electron spins.

(2) The magnitude of the exchange interaction depends upon the electron energy surfaces in valence and conduction bands with the largest contributions coming from states far from the Fermi level. Therefore the experimental interaction constants provide information about these relatively inaccessible regions of the energy bands.

ACKNOWLEDGMENTS

It is a pleasure to thank P. W. Anderson not only for the use of his calculations but also for many stimulating and informative discussions. N. Bloembergen first suggested the appropriateness of an exchange interaction. H. J. Hrostowski kindly provided the high quality samples used in this work.

J. M. M. wishes to thank H. S. Gutowsky for much good advice regarding building of the apparatus. He is also indebted to many colleagues at Bell Telephone Laboratories, particularly F. E. Radcliffe for advice in electronic matters and C. R. Geith for his excellent mechanical design and construction.

[30] P. W. Anderson, J. Phys. Soc. Japan 9, 316 (1954).

Visible Light from a Silicon *p-n* Junction*

ROGER NEWMAN
General Electric Research Laboratory, Schenectady, New York
(Received June 1, 1955)

When low-voltage silicon *p-n* junctions are biased in the reverse direction to breakdown, visible light is emitted from the junction region. The effects of surface treatment on the phenomenon are discussed. Two typical light output *vs* reverse current curves are shown. A typical spectral distribution curve in the 1.8-3.4 ev range is shown. The observations suggest that the light results from a radiative relaxation mechanism involving the high-energy carriers produced in the avalanche breakdown process.

IT has been observed by R. N. Hall and W. E. Burch of this Laboratory that when low-voltage breakdown (<100 volts) silicon *p-n* junctions are biased in the reverse direction to breakdown that a weak yellowish light is emitted from the junction region. This paper describes some studies of this effect. Figure 1 shows a photograph of such a junction taken by its own emitted light. Units obtained from both grown and alloyed junctions in which the breakdown voltage ranged from 10 to 70 volts showed the effect. The data quoted in this report are for units cut from a grown junction with a breakdown voltage of about 40 volts.

When freshly etched, the units showed a hard breakdown characteristic, that is, a small reverse current (e.g., $\leqslant 0.01$ μa at 3 volts for units about 0.05 cm^2 in cross section) that increased slowly with voltage to a sharp break.[1,2] In the breakdown region at current levels less than about 100 ma, the etched units showed small spots of light estimated to be about 10μ or less in diameter. A given spot seemed to appear and reach its maximum brightness over a relatively narrow range of current. Different spots would appear at different current levels. One could cycle through various low current levels and see the different spots appear and disappear reproducibly at characteristic currents. At relatively high current levels, (i.e., greater than about 200 ma or 4 amp/cm^2 for the units being discussed) regions would appear of more or less uniform brightness which often extended several mm along the junction. If no attempt was made to limit the temperature rise in the unit, the brightness was a sensitive function of current. However at a constant current the brightness decreased if the sample was cooled by an air blast, indicating a temperature effect. Hence, relative constancy of temperature was an important factor in any attempt at a quantitative experiment.

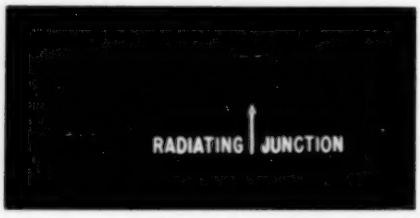


FIG. 1. A photograph of the light emitted from a worked silicon *p-n* junction unit operating in the breakdown region. The junction is the horizontal bow-shaped curve. Current flows vertically across it.

Attempts at controlling the temperature of the units were not successful. An air blast provided insufficient heat exchange. Immersing the sample in various stirred baths of organic liquids (e.g., silicone oil, mineral oil) was sufficient to hold a temperature, but irreversible changes in the units, presumably surface effects, were produced. This was manifested by the nonreproducibility of quantitative measurements of the light output as a function of current.

It was found that when the surface of an etched unit was lightly worked, spots of light would develop at the points where working had taken place. With sufficient working, almost the whole junction region could be made to emit. However, when examined under a microscope it was seen that the light was being emitted from a large number of closely spaced spots. The photograph in Fig. 1 was of such a worked unit. Working the surface consisted, variously, of grinding the surface with various grades of grinding, or polishing powder, or scratching across the surface with a point of stainless steel, tungsten carbide, or quartz. It was found that passing a point under a few grams load across the junction was sufficient to produce the effect, although in many cases the surface showed no indication of scratching. Working, of course, also softened the breakdown characteristic. A single pass of a point was sufficient to increase the reverse current by a factor of 10 or more (e.g., to $\sim$0.1 μa at 3 volts).

In Fig. 2 are shown two plots of reverse current *vs* total light output as measured by an S-4 photomultiplier (EMI 6094). One curve is for a freshly etched unit with junction cross section 1 cm$\times$0.05 cm. The other curve is for the same unit after working it with a quartz point. In both cases, the unit was cooled by an air blast and by conductive cooling through large area Ga-Al paste contacts to copper plates. Below 50 ma, neglibible heating of the silicon occurred.

* A preliminary report of this work was given at the Baltimore American Physical Society Meeting, 1955; Newman, Dash, Hall, and Burch, Phys. Rev. **98**, 1536(A) (1955).
[1] K. G. McKay and K. B. McAfee, Phys. Rev. **91**, 1079 (1954).
[2] K. G. McKay, Phys. Rev. **94**, 877 (1954).

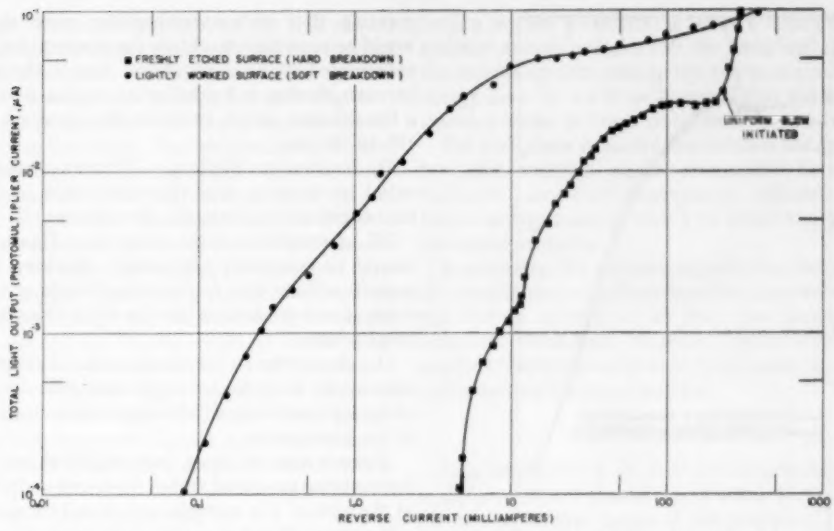


FIG. 2. Two typical reverse current *vs* total light output curves for a silicon p–n junction unit.

The foregoing qualitative observations provide some clue to the interpretation of the curves of Fig. 2. Consider, first, the data for a freshly etched unit. One can suppose that the rapid increase in light output from about 5 to 10 ma corresponds to the development of several bright spots of similar characteristics. Their brightness saturates with increased current corresponding to the break in the curve at about 10 ma. As the current is further increased, a new set of spots start developing in intensity and another rapid increase in light intensity is noted. This process is presumably continued in less discrete steps until by about 100 ma saturation of all the bright spots occurs. Above about 200 ma, a large increase in light output occurs as the relatively large regions of uniform brightness begin to contribute. Ideally one would like to study the characteristics of this uniform brightness region. It is assumed that this effect is associated with the true properties of the junction and not with accidental imperfections in the surface. However, several units were examined using various etching techniques and in no case was a unit obtained which was free from light spots. Consequently it did not appear feasible to obtain a light output *vs* current curve that would relate to one phenomenon alone (i.e., uniform emission) and thus be susceptible to theoretical treatment.

The data of Fig. 2 for the worked unit may be interpreted in terms of a large number of spots, each beginning to contribute its intensity at some different current level. The effects would overlap to give a monotonic appearance to the over-all curve. Curves for units, where the light was emitted over most of the current range from one worked point, resembled the data shown for the worked unit.

It seems pertinent at this point to make some remarks about breakdown in silicon diode units. We have observed that the softness of breakdown and the magnitude of the reverse current can be correlated to intentional imperfections in the surface (e.g., scratches). These, in turn, are manifested as spots of light on the junction when a reverse current in the breakdown range is passed through the unit. It is, therefore, not unreasonable that the spots of light that are observed in etched units may likewise manifest surface imperfections unintentionally produced. Further, the fact that the spots of light appear at different levels of current (or voltage) strongly suggests that a soft breakdown results from breakdown of small patches of material at a spectrum of voltage levels below that required for the breakdown of the bulk material. That is, a spot may be regarded as manifesting a channel for conduction of lower impedance than the main body of material. This suggests the possibility that the reverse currents that are usually measured in silicon units at voltages well below breakdown may be dominated by surface imperfections. These ideas are, of course, not new or particularly surprising. However, it would seem that the present results constitute direct evidence for this view of soft breakdown effects.

It had been hoped that by irradiating the junction with infrared radiation and thus altering the reverse current at constant voltage that an effect could be observed which would definitely associate the light emission with the avalanche breakdown in the silicon.[1] Our experiment was inconclusive in that over the current range where observable changes in reverse current could be obtained no light was detectable.

Figure 3 shows the spectral distribution of the light

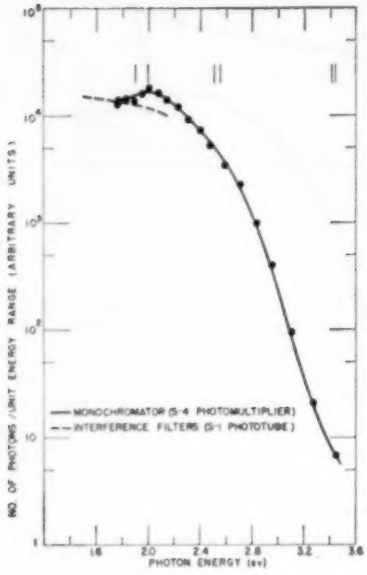

NO. OF PHOTONS/UNIT ENERGY RANGE (ARBITRARY UNITS)
MONOCHROMATOR (S-4 PHOTOMULTIPLIER)
INTERFERENCE FILTERS (S-1 PHOTOTUBE)
PHOTON ENERGY (ev)

FIG. 3. Spectral distribution curve of the light emitted from a silicon p-n junction operated in the breakdown region.

emitted by an etched unit operated in the region of uniform brightness and cooled with an air blast. The measurements were made using a low dispersion quartz monochromator with a photomultiplier detector. The source light was chopped at 450 cps and a Baird phase-sensitive amplifier was used in conjunction with a recording potentiometer. In order to get a sufficient signal/noise ratio it was necessary to operate at an overall band width of about 0.2 cps. The data shown are corrected for the variation in spectral slit width of the monchromator and for the variation in the spectral sensitivity of the photomultiplier. The spectral distribution curve for a worked unit was identical with that shown to within our accuracy of measurement.

Attempts were made to extend the range of the spectral measurements toward lower energies. Vacuum and gas S-1 phototubes and a cooled PbS cell were tried with generally unsatisfactory results. Some data were obtained using interference filters and a cooled vacuum phototube in an electrometer circuit capable of detecting a photocurrent of 10^{-15} amp. These are indicated by the dotted curve of Fig. 3.

The following represent some hypotheses concerning the origin of the effect and the evidence that bears upon them:

I. *Local heating of the silicon.*—Considerations of thermal flow would seem to eliminate this as a possibility. With reasonable estimates of power input densities and estimates of thermal loss, calculations indicate that temperatures characteristic of the emitted light, that is ~2000°K, would not be attainable. Further, even granting that such a temperature were attainable, would be impossible to obtain the temperature gradient necessitated by the observations. That is, the boundary between glowing and nonglowing regions is defined to a few microns, which would imply a gradient of order 10^6–10^7°K/cm.

II. *External gas discharge.*—This was readily eliminated by showing that the effect could be obtained unchanged in a vacuum of a few microns.

III. *A breakdown in the oxide film.*—This possibility cannot be completely discounted. However, it would appear unlikely that a breakdown would occur in the oxide at as low a field as that for which the silicon itself breaks down.

IV. *Breakdown in the silicon itself.*—This possibility seems most likely from the apparently close correlation of the light emission with the breakdown characteristics of the silicon units.

There is some evidence that would indicate that the light is being generated within the silicon and not simply at the surface. For example, occasionally a spot of light will appear colored asymmetrically; that is one side of the spot will appear yellow and the other side of the spot will appear orange or red. This is understandable if the breakdown process at the spot penetrated the silicon nonuniformly. The absorption coefficients are such that the penetration depth for yellow light is ~1.5μ and ~3μ for red light.[3] Occasionally even blue tinted spots could be formed but these would either degrade to yellow or disappear, consistent with a very small penetration depth for blue light (i.e., ~0.3μ) and an unstable surface. McKay, using broad area diffused junctions, has also found color variations in the light consistent with emission from the bulk.[4]

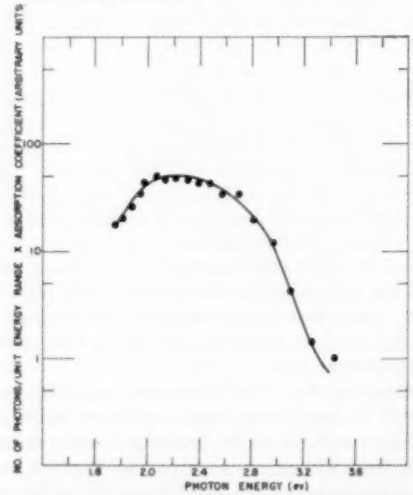


FIG. 4. Data of Fig. 3 corrected for self-absorption of the silicon.

[3] Dash, Newman, and Taft, Phys. Rev. **98**, 1192(A) (1955).
[4] K. G. McKay and A. G. Chynoweth, Bull. Am. Phys. Soc. **30**, No. 4, 22 (1955).

If the light emission is coming from the bulk silicon, then the spectral distribution curve should be corrected for the self-absorption of the sample. The spectral distribution curve so corrected is indicated in Fig. 4.

If the light is identified with the avalanche breakdown process in the silicon, then two possible processes suggest themselves. The first process involves the radiative recombination of the high-energy electrons and holes produced in the junction region during breakdown. The radiation produced by this process would presumably have a low-energy threshold at about the band gap (~1 ev) of silicon. Since the probability of carriers having a given energy will decrease with increasing energy the spectral distribution would, of course, show a tailing off at the high-energy end as well.[5] The spectral distribution for this case might resemble the curve of Fig. 4. The results of Wolff's calculation[5] of the carrier distribution function would appear consistent with the observed spectral distribution curve. It would be expected that the light output would increase as the square of the current for this mechanism.

The second process involves an intraband relaxation That is, a high-energy carrier could lose its energy by radiation and drop into a lower level in its own band. Such a process would also produce a tailing off of the emission at high energies (see the foregoing). However, the radiation would presumably, not have a definite low-energy limit. It would be expected that the total light emission would be linear in current.

For both these radiation mechanisms the decay time for light emission would presumably be of order 0.1−0.01 μsec.[2] Our measurement indicated a decay time that was 4 μsec or less. This result was limited by the instrumentation.

In summary, the evidence suggests that the light emission results from radiative relaxation processes involving high-energy carriers in or near the barrier during avalanche breakdown. However, further study will be required before this conclusion can be made rigorous and a detailed mechanism elucidated.

[5] P. Wolff, Phys. Rev. **95**, 1415 (1954).

ACKNOWLEDGMENTS

I am indebted to R. N. Hall for bringing this problem to my attention. I should like to thank W. C. Dash for his help at various stages of the project. E. O. Kane suggested the intraband relaxation mechanism and was very helpful in discussing various phases of this work. I am also indebted to K. G. McKay for his comments on this manuscript.

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Stark Effect in Rapidly Varying Fields*

S. H. Autler† and C. H. Townes
Columbia University, New York, New York

(Received May 31, 1955)

A method is developed for calculating the effects of a strong oscillating field on two states of a quantum-mechanical system which are connected by a matrix element of the field. Explicit approximate solutions are obtained for a variety of special cases, and the results of numerical computations are given for others. The effect of an rf field on the $J=2\rightarrow1$ l-type doublet microwave absorption lines of OCS has been studied in particular both experimentally and theoretically. Each line was observed to split into two components when the frequency of the rf field was near 12.78 Mc or 38.28 Mc, which are the frequencies separating the $J=1$ and $J=2$ pairs of levels, respectively. By measuring the rf frequency, ν_0, at which the microwave lines are split into two equally intense components, one may determine the separation between the energy levels. The measured value of ν_0 depends upon the intensity of the rf field and the form of this dependence has been calculated and found to be in good agreement with the experimental results.

I. INTRODUCTION

1.1 Outline of the Problem

A RELATIVELY weak perturbation varying sinusoidally in time may affect a physical system by causing an occasional transition between quantum states. These transitions are accompanied by the absorption or emission of photons and may be observed spectroscopically. However, if the perturbation is strong enough, transitions are rapidly induced and a variety of other observable phenomena can occur. We shall discuss some of these in the case of a system where relaxation processes are negligible, i.e., where the effects of the sinusoidal perturbation are much more important than those due to processes which dissipate energy. In particular, an experiment will be described which involves the simultaneous effects of two electromagnetic fields on the molecules of gaseous OCS. One field is much stronger than the other and is in the radio-frequency range; the weaker field is in the microwave

* Work supported jointly by the Department of the Army (Signal Corps), the Department of the Navy (Office of Naval Research), and the Department of the Air Force (Air Research and Development Command).

† Present address: Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts.

region and is transmitted through a wave guide containing the gas while absorption lines are observed by the usual techniques of microwave spectroscopy. At certain frequencies the rf field induces rapid transitions between the molecular states, which affects the absorption of the microwaves. The absorption of the rf field, which is not directly observable can thus be indirectly studied by microwave spectroscopy.

After a statement of definitions and a short introductory theory, some of the previously published work related to our subject is discussed. Then a general treatment is developed for the effects of a sinusoidally varying field of any strength and frequency on a simple quantum-mechanical system for which only two stationary states need be considered. The general solution is expressed in terms of infinite continued fractions, but simple approximations are given which are valid over certain ranges of field strength and frequency; for other ranges numerical results are given.

The case of "resonant modulation" where the frequency of the field coincides with an internal frequency of the system is of particular interest. Although our theoretical treatment was developed primarily for this case, and is compared in some detail with experimental results, it is applicable to a wider range of conditions including some found in molecular beam experiments.

1.2 Preliminary Theory and Definitions

The discussion of the effects of an oscillating field will be simplified by assuming that the wave function of the system may be expressed as a linear combination of just two of its unperturbed eigenfunctions. Thus,

$$\Psi = T_a(t)U_a + T_b(t)U_b, \tag{1}$$

where U_a and U_b are the space-dependent parts of these two eigenfunctions. The states are assumed to be nondegenerate and to have energies W_a and W_b with $W_b > W_a$. Relaxation processes will be completely ignored, so the energy levels are treated as perfectly sharp.

We may expect Eq. (1) to be a valid approximation in a number of cases. If the frequency of the perturbing field is close to $(W_b - W_a)/h$, it has a much greater effect on these two states than on others, which may then be ignored. This is also permissible for any frequency if W_a and W_b are much closer to each other than to any other states; this requirement is satisfied by the states of the OCS molecule which were studied in our experiment. Later a microwave field will be allowed to induce occasional transitions to a third state well removed in energy, but if this field is weak it has a relatively small effect on the states under study.

The time-dependent wave equation is

$$i\hbar\dot{\Psi} = [H^0 - \boldsymbol{\mu}\cdot\mathbf{E}\cos\omega t]\Psi, \tag{2}$$

where H^0 = unperturbed Hamiltonian containing all internal interactions, $\boldsymbol{\mu}$ = permanent electric dipole moment of molecule, $\mathbf{E}\cos\omega t$ = applied electric field. Substituting (1) into (2) gives

$$\begin{aligned} i\hbar\dot{T}_a &= W_aT_a + 2\hbar\beta_{ab}\cos\omega t\, T_b, \\ i\hbar\dot{T}_b &= W_bT_b + 2\hbar\beta_{ba}\cos\omega t\, T_a, \end{aligned} \tag{3}$$

where

$$\beta_{ab} = \frac{\langle a|\boldsymbol{\mu}\cdot\mathbf{E}|b\rangle}{2\hbar}, \quad \beta_{ab} = \beta_{ba}$$

$$H^0U_a = W_aU_a, \quad H^0U_b = W_bU_b.$$

β_{aa} and β_{bb} equal zero if states a and b are nondegenerate. β_{ab} may be assumed real without loss of generality as any phase factor may be absorbed into T_a and T_b.

In spite of their rather simple appearance, these equations have not been solved in terms of tabulated functions. The complicated form of the equation

$$0 = \frac{\partial^2 T_a}{\partial t^2} + \left[\frac{i}{\hbar}(W_a+W_b) + \omega\tan\omega t\right]\frac{\partial T_a}{\partial t} + \left[4\beta_{ab}^2\cos^2\omega t - \frac{W_aW_b}{\hbar^2} + \frac{iW_a}{\hbar}\omega\tan\omega t\right]T_a$$

obtained by eliminating T_b from (3) indicates why this is so. One must therefore obtain approximate solutions for various values of the parameters β_{ab} and ω.

The ranges of β_{ab} and ω will be denoted as follows:

Field strength		Frequency	
Weak:	$\beta_{ab} \ll \omega_{ab}$	Low:	$\omega \ll \omega_{ab}$
Intermediate:	$\beta_{ab} \approx \omega_{ab}$	Resonant:	$\omega \approx \omega_{ab}$
Strong:	$\beta_{ab} \gg \omega_{ab}$	High:	$\omega \gg \omega_{ab}$

$\omega_{ab} = (W_b - W_a)/h.$

II. REVIEW OF RELATED WORK

2.1 Resonant Modulation

The authors have previously reported results of an experiment[1] in which a radio-frequency field was used to produce transitions between two energy levels of the molecules of OCS gas and thus modulate the probability of finding a molecule in either state. The gas was kept in a wave guide and a microwave absorption line involving one of the levels was observed while the rf field was applied across the wave guide. When the frequency of the rf field was made resonant with the separation between the two levels, the microwave absorption line appeared to split into two components of nearly equal intensity. Although the possibility of measuring energy differences by this method was established, there were certain discrepancies between observations and theory. In this paper, more precise experimental data and more detailed theory are reported and better agreement is obtained.

Time-dependent perturbation calculations are carried into a somewhat broader range of frequencies and field

[1] S. H. Autler and C. H. Townes, Columbia Radiation Laboratory Reports, June 30, 1949 and September 30, 1949 (unpublished). S. H. Autler and C. H. Townes, Phys. Rev. **78**, 340 (1950).

strengths in this paper than has previously been reported. Some previous results are reviewed below in order to show their relationship to the present work.

2.2 Extremely Weak Fields

In discussing the effect of a resonant field in inducing transitions between two stationary states, it may be assumed that the system is originally known to be in one of these states and that the probability of finding it in the other state remains small.[2] This gives a transition probability which increases as t^2. It is then usually assumed that the final state can be any one of a very closely spaced group of states, or that the radiation consists of a band of frequencies. This leads to the familiar result in which the transition probability grows linearly in time at a rate proportional to the square of the field strength until a collision or some other relaxation mechanism terminates the process.

2.3 Rotating Fields

For fields which are somewhat stronger, though still weak according to the definition adopted above, it is no longer possible to neglect the probability of finding the system in the second state. This problem has been discussed by Majorana and by Rabi[3] for the case of a magnetic dipole in a rotating magnetic field. For a spin of $\frac{1}{2}$, Rabi obtained the equations

$$\begin{aligned} i\hbar\dot{C}_{+\frac{1}{2}} &= \frac{g\mu_0 H_z}{2}C_{+\frac{1}{2}} + \frac{g\mu_0 H_\perp}{2}e^{-i\omega t}C_{-\frac{1}{2}}, \\ i\hbar\dot{C}_{-\frac{1}{2}} &= -\frac{g\mu_0 H_z}{2}C_{-\frac{1}{2}} + \frac{g\mu_0 H_\perp}{2}e^{+i\omega t}C_{+\frac{1}{2}}, \end{aligned} \tag{4}$$

where $H_\perp$ = component of rotating field which lies $\perp$ to H_z, H_z = average value of magnetic field $\mathbf{H}$, g = Landé g-factor of particle, μ_0 = Bohr magneton. $C_{+\frac{1}{2}}$ and $C_{-\frac{1}{2}}$ are the probability amplitudes of the two spin orientations.

The form of these equations differs from that of Eqs. (3) only in the replacement of $\cos\omega t$ by exponentials. This difference, however, allows Eqs. (4) to be solved exactly.

The result obtained is that the probability of finding the dipole in the $-\frac{1}{2}$ state, assuming it starts in the $+\frac{1}{2}$ state, is:

$$P(\tfrac{1}{2}, -\tfrac{1}{2}) = \frac{(g\mu_0 H_\perp)^2}{\hbar^2(\omega-\omega_0)^2 + (g\mu_0 H_\perp)^2} \times \sin^2\left\{\frac{t}{2}\left[(\omega-\omega_0)^2 + \left(\frac{g\mu_0 H_\perp}{\hbar}\right)^2\right]^{\frac{1}{2}}\right\}, \tag{5}$$

where $\omega_0 = g\mu_0 H_z/\hbar$ is the Larmor angular frequency of the dipole in the field.

The particle oscillates sinusoidally between the two states with an amplitude which is equal to one when $\omega = \omega_0$.

[2] L. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1949), Chap. X.
[3] I. I. Rabi, Phys. Rev. **51**, 652 (1932).

Torrey[4] calculated this in a somewhat different manner in connection with an investigation of the line shape of molecular beam radio-frequency resonances. We shall use an approach similar to Torrey's, whose result is obtained as the first approximate solution of our general equations in Sec. 3.3.

2.4 Weak Oscillating Field

Equation (5), which describes precisely the effect of any rotating field, is also a good approximation for an oscillating field if that field is weak and its frequency near resonance. This is to be expected since a field varying as $\cos\omega t$ can be separated into two rotating fields, $\frac{1}{2}e^{+i\omega t}$ and $\frac{1}{2}e^{-i\omega t}$. One of these can be thought of as rotating with the precessing particle and the other oppositely. A small oppositely rotating field will have little effect compared to the resonant one. Bloch and Siegert[5] made an approximate calculation of this effect for the case of a particle of spin $\frac{1}{2}$ in a polarizing magnetic field by expressing the equivalent of Eq. (3) as an integral equation which was then expanded in a series. They showed that the first effect of the counter-rotating field is to increase the resonant frequency by the fractional amount $\frac{1}{16}(H_\perp/H_z)^2$. In our notation this can then be written

$$\omega_0/\omega_{00} = 1 + (\beta_{00}^2/\omega_{00}^2),$$

where ω_0 is the observed resonant frequency. This result is discussed in Sec. 4.2.

2.5 Double Modulation

The simultaneous use of radio-frequency and ultraviolet radiation to investigate the structure of atomic energy levels has been described by Brossel, Kastler, and Bitter.[6] After exciting mercury vapor atoms by resonance radiation, they ingeniously employ the rf field to alter the population of the atoms in the excited state and then observe the change in polarization of the spontaneously emitted ultraviolet radiation. They obtained a radio-frequency resonance and splitting similar to that described in Sec. 2.1. The theory given for "resonant modulation"[1] is in fact applicable to Brossel and Bitter's case. However, Pryce[7] has independently calculated the splitting which occurs in this experiment. Brossel and Bitter also obtain a shift in the resonance approximately equal to that calculated by Bloch and Siegert.[5]

2.6 Low-Frequency Perturbation

If the frequency of the applied field is sufficiently low, its effect can be treated quasi-statically, i.e., the solution can be regarded at any instant as being the steady-state solution appropriate for the instantaneous field. The instantaneous value of the field appears as a parameter in the solution, and as the field varies, the solution is presumed to follow along.

[4] H. C. Torrey, Phys. Rev. **59**, 293 (1941).
[5] F. Bloch and A. Siegert, Phys. Rev. **57**, 522 (1940).
[6] J. Brossel and A. Kastler, Compt. rend. **229**, 1213 (1949); J. Brossel and F. Bitter, Phys. Rev. **86**, 308 (1952).
[7] M. H. L. Pryce, Phys. Rev. **77**, 136 (1950).

Townes and Merritt[8] calculated and observed the second-order Stark effect due to fields of several hundred kilocycles, and also calculated the first order case. They found that when ω is larger than the width of the absorption line each Stark component breaks up into lines spaced by the angular frequency ω, with amplitudes depending upon the field strength. In second-order Stark effect each energy level breaks up into a number of components with angular frequencies displaced from its unperturbed value by $\frac{1}{2}aE^2\pm 2n\omega$, where n is any integer, a is a constant, and E the field strength. Relative intensities of the components are given by $J_n^2(aE^2/2)$. The agreement between theory and experiment was found to be excellent. A derivation of the above result is given in Sec. 3.5.

Since the present work ignores the effects of collisions, the case where ω is comparable to the line width requires another treatment. This has been given by Karplus,[9] who used the methods of quantum statistics and treated in detail several cases where the molecular resonances are shifted by fields with particular wave forms.

III. THEORY

3.1 General Formulas

A solution is desired for T_a and T_b in Eq. (3). Use will be made of Floquet's theorem[10] which shows that T_a and T_b have the form

$$T_a=e^{i\lambda t}\sum_{n=-\infty}^{+\infty}A_n e^{-in\omega t},\quad T_b=e^{i\lambda t}\sum_{n=-\infty}^{+\infty}B_n e^{-in\omega t},\quad (6)$$

where n is any positive or negative integer or zero. If (6) is substituted into (3) after expressing $\cos\omega t$ in exponentials, and terms with the same exponents are equated, one obtains

$$\begin{aligned}(\lambda+\omega_a-n\omega)A_n&=-\beta_{ab}B_{n-1}-\beta_{ab}B_{n+1},\\(\lambda+\omega_b-n\omega)B_n&=-\beta_{ab}A_{n-1}-\beta_{ab}A_{n+1},\end{aligned}\quad (7)$$

where $\omega_a=W_a/\hbar$, $\omega_b=W_b/\hbar$. This set of equations can be separated into two completely independent sets as follows:

$$\begin{aligned}[L_a-k\omega/\beta_{ab}]A_k&=-B_{k-1}-B_{k+1},\\[L_a+\omega_{ab}/\beta_{ab}-k\omega/\beta_{ab}]B_l&=-A_{l-1}-A_{l+1},\end{aligned}\quad (8)$$

with all odd A's and even B's being zero. Also

$$\begin{aligned}[L_b-\omega_{ab}/\beta_{ab}-l\omega/\beta_{ab}]A_l&=-B_{l-1}-B_{l+1},\\[L_b-k\omega/\beta_{ab}]B_k&=-A_{k-1}-A_{k+1},\end{aligned}\quad (9)$$

with all odd B's and even A's equal to zero. Here

$$L_a\equiv\frac{\lambda+\omega_a}{\beta_{ab}},\quad L_b\equiv\frac{\lambda+\omega_b}{\beta_{ab}},$$

and k runs through all positive and negative even integers and zero; l runs through all positive and negative odd integers.

Since Eqs. (8) and (9) are independent, a complete solution may be obtained by adding any solution of (8) to any solution of (9).

We will first concentrate on (8), an infinite set of linear homogeneous equations with an infinite number of unknown quantities (the A's, B's, and λ) to be determined. One could set up the infinite secular determinant of the coefficients and attempt to solve it. Instead, the method to be used here is to express the ratios of the A's and B's as infinite continued fractions and then obtain an equation for λ involving them. This method of attack arises naturally and the continued fraction expressions are fairly amenable to algebraic and numerical calculations to any accuracy.

In Appendix 3.1a, it is shown that if one assumes that the amplitude coefficients (A's and B's) become negligible for $|k|$ and $|l|$ sufficiently large, then

$$\frac{A_{\pm k}}{B_{\pm(k-1)}}=\cfrac{-1}{L_a\mp k\frac{\omega_{ab}}{\beta_{ab}}\left(\frac{\omega}{\omega_{ab}}\right)-\cfrac{1}{L_a+\frac{\omega_{ab}}{\beta_{ab}}\mp(k+1)\frac{\omega_{ab}}{\beta_{ab}}\left(\frac{\omega}{\omega_{ab}}\right)-\cfrac{1}{L_a\mp(k+2)\frac{\omega_{ab}}{\beta_{ab}}\left(\frac{\omega}{\omega_{ab}}\right)-\cfrac{1}{L_a+\frac{\omega_{ab}}{\beta_{ab}}\mp(k+3)\frac{\omega_{ab}}{\beta_{ab}}\left(\frac{\omega}{\omega_{ab}}\right)-\cdots}}}},\quad (10)$$

$$\frac{B_{\pm l}}{A_{\pm(l-1)}}=\cfrac{-1}{L_a+\frac{\omega_{ab}}{\beta_{ab}}\mp l\frac{\omega_{ab}}{\beta_{ab}}\left(\frac{\omega}{\omega_{ab}}\right)-\cfrac{1}{L_a\mp(l+1)\frac{\omega_{ab}}{\beta_{ab}}\left(\frac{\omega}{\omega_{ab}}\right)-\cfrac{1}{L_a+\frac{\omega_{ab}}{\beta_{ab}}\mp(l+2)\frac{\omega_{ab}}{\beta_{ab}}\left(\frac{\omega}{\omega_{ab}}\right)-\cfrac{1}{L_a\mp(l+3)\frac{\omega_{ab}}{\beta_{ab}}\left(\frac{\omega}{\omega_{ab}}\right)-\cdots}}}}.$$

[8] C. H. Townes and F. R. Merritt, Phys. Rev. **72**, 1266 (1947).
[9] R. Karplus, Phys. Rev. **73**, 1027 (1948).
[10] H. Margenau and G. M. Murphy, *Mathematics of Physics and Chemistry* (D. Van Nostrand Company, Inc., New York, 1947), p. 80.

In reading the above expressions, either the upper or lower sign should be used throughout.

In Eq. (10), the ratios of successive amplitude coefficients are expressed as functions of ω_{ab}/β_{ab} and ω/ω_{ab}, two dimensionless quantities which are natural measures of the intensity and frequency of the applied field and completely determine the problem. L_a is defined as $(\lambda+\omega_a)/\beta_{ab}$ and in Eq. (11) below is given as a function of ω_{ab}/β_{ab} and ω/ω_{ab}. Equations (10) then may be regarded as expressing successive amplitude coefficients (A's and B's) as functions of the physical variables. For example, by letting l equal one, an expression for B_{+1}/A_0 is obtained. Similarly, $A_{\pm 2}/B_{\pm 1}$, $B_{\pm 3}/A_{\pm 2}$, etc., can all be expressed as continued fractions, and the ratio of any of the A's and B's to A_0 written as a product of these quantities. A_0 is one of the arbitrary constants which appear in a general solution of a second order differential equation, and may be evaluated by applying normalization and initial conditions to ψ.

Besides being necessary for calculating the amplitude coefficients, L_a has a physical interpretation which will be discussed at the end of Sec. 3.3.

By setting $k=0$ in Eq. (8) and then making use of Eq. (10), we obtain

$$L_a=\cfrac{1}{L_a+\cfrac{\omega_{ab}}{\beta_{ab}}\left(1-\cfrac{\omega}{\omega_{ab}}\right)-\cfrac{1}{L_a-2\left(\cfrac{\omega_{ab}}{\beta_{ab}}\right)\left(\cfrac{\omega}{\omega_{ab}}\right)-\cfrac{1}{L_a+\cfrac{\omega_{ab}}{\beta_{ab}}\left(1-\cfrac{3\omega}{\beta_{ab}}\right)-\cdots}}}+\cfrac{1}{L_a+\cfrac{\omega_{ab}}{\beta_{ab}}\left(1+\cfrac{\omega}{\omega_{ab}}\right)-\cfrac{1}{L_a+2\left(\cfrac{\omega_{ab}}{\beta_{ab}}\right)\left(\cfrac{\omega}{\omega_{ab}}\right)-\cdots}} \tag{11}$$

Equations (10) and (11) are general expressions which, at least in principle, can be used to determine L_a and then the amplitudes for all values of ω_{ab}/β_{ab} and ω/ω_{ab}. To actually solve (11) for L_a, one must in general retain only a finite number of the quotients in each continued fraction, but by retaining a sufficient number L_a can be evaluated to any desired accuracy. Eventually, the denominators are dominated by the terms of the form $n\omega/\beta_{ab}$ and the quotients approach zero; which is a necessary condition of the fractions are to converge. The larger ω/β_{ab}, the more rapid the convergence, so it will be necessary to retain fewer quotients for high frequencies and weak fields.

No proof has been obtained that the continued fractions always converge. However, it is shown in the Appendix 3.1b that when $\omega=0$, assuming convergence leads to the correct steady-state solution. That this is so in the case when one expects least rapid convergence indicates that the fractions probably converge in all other cases.

If only a few quotients need to be retained, Eq. (11) can be expressed as an algebraic equation of low degree and solved either exactly or approximately by familiar methods. This will be done for weak fields. For stronger fields where more quotients must be kept, it is easier to work numerically with the continued fractions themselves, and solutions for L_a have been obtained numerically for a wide range of frequencies and for fairly strong fields.

The equation obtained by retaining all terms and expressing Eq. (11) in powers of L_a, would be of infinite degree and have an infinite number of solutions. Fortunately, because of the periodic properties of the original differential equation, all these solutions are related in a simple manner, so that if one is known all the others may be readily obtained. In Appendix 3.1c, the following relations are proved:

Assume L_a, A_0, A_2, $\cdots$, B_1, B_3, $\cdots$ satisfy Eqs. (8). Then L_a', A_0', A_2', $\cdots B_1'$, B_3', $\cdots$ give another solution if

$$L_a'=-L_a-\frac{\omega_{ab}-(2m+1)\omega}{\beta_{ab}}, \tag{12}$$

and

$$B_l'=A_{2m+1-l}, \quad A_k'=-B_{2m+1-k}, \tag{13}$$

where m is any positive or negative integer.

Another set of solutions of (8) is

$$\begin{aligned} L_a''&=L_a+2m(\omega/\beta_{ab}),\\ A_k''&=A_{k-2m}, \quad B_l''=B_{l-2m}. \end{aligned} \tag{14}$$

There are other solutions which satisfy (9) rather than (8). In Appendix 3.1c, the following simple relationship between the solutions of (8) and (9) is proved: If L_a, $A_k/A_0\cdots$, $B_l/A_0\cdots$ define any solution of (8), then L_b, $A_l/B_0\cdots$, $B_k/B_0\cdots$ define a solution of (9), if

$$L_b=-L_a;\ A_l/B_0=-B_{-l}/A_0,\ B_k/B_0=+A_{-k}/A_0. \tag{15}$$

The solutions of (8) contain only even A's and odd B's while the reverse is true for (9).

A complete solution is obtained by adding any solution of (8) to any solution of (9). It is quite arbitrary which of the infinite number of solutions is used in each case, for they all contain the same physical information; the various solutions corresponding essentially to a relabeling of the amplitude coefficients. However, it is necessary to understand the nature of the solutions when actually making calculations to avoid errors which might arise by skipping from one solution to another without knowing it.

A complete solution is obtained by combining Eqs. (1) and (6).

$$\psi \equiv U_a\left\{A_0 e^{i\lambda_a t}\sum_{\substack{k=-\infty\\ \text{even}}}^{+\infty}\frac{A_k}{A_0}e^{-ik\omega t} + B_0 e^{i\lambda_b t}\sum_{\substack{l=-\infty\\ \text{odd}}}^{+\infty}\frac{A_l}{B_0}e^{-il\omega t}\right\} + U_b\left\{A_0 e^{i\lambda_a t}\sum_{\substack{l=-\infty\\ \text{odd}}}^{+\infty}\frac{B_l}{A_0}e^{-il\omega t} + B_0 e^{i\lambda_b t}\sum_{\substack{k=-\infty\\ \text{even}}}^{+\infty}\frac{B_k}{B_0}e^{-ik\omega t}\right\}. \quad (16)$$

Here $\lambda_a, A_k\cdots, B\cdots$ may form any solution of (8). λ_a is equal to $\beta_{ab}L_a-\omega_a$, where L_a is any solution of (11). Also

$$\frac{A_{\pm k}}{A_0}=\frac{A_{\pm k}}{B_{\pm(k-1)}}\cdot\frac{B_{\pm(k-1)}}{A_{\pm(k-2)}}\cdots\frac{B_{\pm 1}}{A_0};$$

$$\frac{B_{\pm l}}{A_0}=\frac{B_{\pm l}}{A_{\pm(l-1)}}\cdot\frac{A_{\pm(l-1)}}{B_{\pm(l-2)}}\cdots\frac{B_{\pm 1}}{A_0};$$

where the factors of the form $A_{\pm k}/B_{\pm(k-1)}$ and $B_{\pm l}/A_{\pm(l-1)}$ are expressed as continued fractions by (10). Similarly, $\lambda_b, A_l\cdots, B_k\cdots$ form a solution of (9).

The number of independent quantities in Eq. (16) may be reduced by first choosing a solution for L_a and then using (15) to derive the solution of (9) to be used. Define

$$\lambda_a=\beta_{ab}L_a-\omega_a, \quad \lambda_b=\beta_{ab}L_b-\omega_b,$$

and set

$$L_b=-L_a.$$

Then

$$\lambda_b=-\lambda_a-\omega_a-\omega_b.$$

Also,

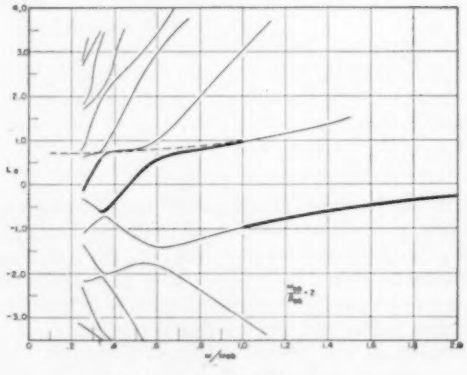
$$\frac{A_l}{B_0}=-\frac{B_{-l}}{A_0}, \quad \frac{B_k}{B_0}=+\frac{A_{-k}}{A_0}.$$

Equation (16) can then be written

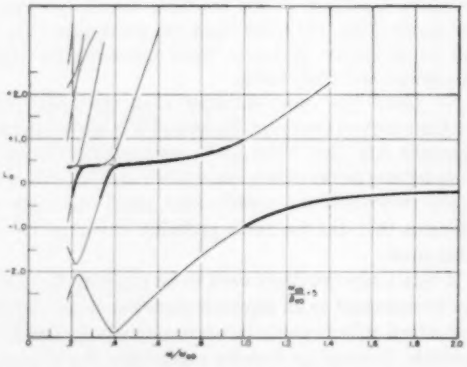
$$\psi=U_a e^{-\frac{1}{2}i(\omega_a+\omega_b)t}\left\{|A_0|e^{i[\frac{1}{2}(\omega_a+\omega_b)+\lambda_a]t}\sum_{\substack{k=-\infty\\ \text{even}}}^{+\infty}\frac{A_k}{A_0}e^{-ik\omega t} - |B_0|e^{-i[\frac{1}{2}(\omega_a+\omega_b)+\lambda_a]t-i\theta}\sum_{\substack{l=-\infty\\ \text{odd}}}^{+\infty}\frac{B_l}{A_0}e^{+il\omega t}\right\} - U_b e^{-\frac{1}{2}i(\omega_a+\omega_b)t}\left\{|A_0|e^{i[\frac{1}{2}(\omega_a+\omega_b)+\lambda_a]t}\sum_{l=-\infty}^{+\infty}\frac{B_l}{A_0}e^{-il\omega t} + |B_0|e^{-i[\frac{1}{2}(\omega_a+\omega_b)+\lambda_a]t-i\theta}\sum_{k=-\infty}^{+\infty}\frac{A_k}{A_0}e^{ik\omega t}\right\}, \quad (17)$$

where

$$e^{i\theta}\equiv\frac{B_0}{|B_0|}\Big/\frac{A_0}{|A_0|}$$



(a) Strong applied field, $\omega_{ab}/\beta_{ab}=2$.



(b) Weaker applied field, $\omega_{ab}/\beta_{ab}=5$.

FIG. 1. The numerically computed solutions of Eq. (11) for L_a as a function of the frequency of the applied field. The darkened curves show the preferred solution in each range of ω/ω_{ab}. The dashed curves are plots of the approximation for L_a given by Eq. (33)

depends upon the relative phase of A_0 and B_0. A_0 and B_0 must be chosen so as to normalize ψ and satisfy the initial conditions. λ_a, $A_k/A_0\cdots$, $B_l/B_0\cdots$ may still form any one of the possible solutions of (8) and the particular one chosen is chiefly a matter of convenience.

In Fig. 1 the solutions of L_a are plotted as a function of the frequency for a number of field strengths. It is based on numerical computations which are discussed at the end of Sec. 3.3.

Although any of the possible solutions for L_a could be used in Eqs. (10) and (17) to give a consistent result, one choice is most natural. This is the solution which remains finite when the field intensity and, therefore, β_{ab} approaches zero. At any value of ω, except for certain discrete points, there is only one such solution; if L_a is this solution, it can be seen by Eqs. (12) and (14) that the other solutions L_a' and L_a'' become infinite as $\beta_{ab}\to 0$. When $\omega/\omega_{ab}=1, \frac{1}{2}, \frac{1}{3}, \cdots$ there are two solutions which remain finite. At these points we will choose the one which is used at slightly larger values of ω. The darkened curves in Fig. 1 are the values of L_a which will be used.

Since $\lambda_a+\omega_a=\beta_{ab}L_a$, λ_a approaches $-\omega_a$ for very small field strength. Equation (10) shows that all components except A_0 and B_0 then become zero, so that Eq. (17) reduces to the unperturbed solution

$$\psi=A_0e^{-i\omega_a t}+B_0e^{-i\omega_b t}.$$

Our choice of L_a thus insures that for weak enough fields A_0 will be the dominant U_a component, except near the special points $\omega/\omega_{ab}=1, \frac{1}{2}, \frac{1}{3}, \cdots$ where another component may be greater.

Up to now, we have concentrated upon obtaining a general solution of the basic differential equations which requires evaluating λ and the A's and B's in Eq. (6). In the next section, we will show how these quantities may be used directly to calculate the effects of the rf field on an observed microwave absorption line.

In molecular beam experiments the values of interest are the probabilities of finding the molecule in either state at any given time. These are equal to $|T_a|^2$ and $|T_b|^2$ in Eq. (1). From (17)

$$\begin{aligned}|T_a|^2=&|A_0|^2\sum_{k=-\infty}^{+\infty}\left(\frac{A_k}{A_0}\right)^2+|B_0|^2\sum_{l=-\infty}^{+\infty}\left(\frac{B_{-l}}{A_0}\right)^2\\&+2\sum_{\substack{k'=-\infty\\ \text{even}}}^{+\infty}\left[|A_0|^2\sum_{k=-\infty}^{+\infty}\left(\frac{A_k}{A_0}\right)\left(\frac{A_{k-k'}}{A_0}\right)\right.\\&\left.+|B_0|^2\sum_{l=-\infty}^{+\infty}\left(\frac{B_{-l}}{A_0}\right)\left(\frac{B_{-l-k'}}{A_0}\right)\right]\cos k'\omega t\\&-2|A_0||B_0|\sum_{k=-\infty}^{+\infty}\sum_{l=-\infty}^{+\infty}\frac{A_k}{A_0}\frac{B_{-k-l}}{A_0}\\&\times\cos[(\omega_a+\omega_b+2\lambda_a+l\omega)t+\theta],\end{aligned}\quad(18)$$

$$\begin{aligned}|T_b|^2=&|A_0|^2\sum_{l=-\infty}^{+\infty}\left(\frac{B_l}{A_0}\right)^2+|B_0|^2\sum\left(\frac{A_{-k}}{A_0}\right)^2\\&\times\sum_{k'=-\infty}^{+\infty}\left\{|A_0|^2\sum_{l=-\infty}^{+\infty}\frac{B_l}{A_0}\frac{B_{l-k'}}{A_0}\right.\\&\left.+|B_0|^2\sum_{k=-\infty}^{+\infty}\frac{A_{+k}}{A_0}\frac{A_{k-k'}}{A_0}\right\}\cos k'\omega t\\&+2|A_0||B_0|\sum_{l=-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty}\frac{A_{-k}}{A_0}\frac{B_{k-l}}{A_0}\\&\times\cos[(\omega_a+\omega_b+2\lambda_a+l\omega)t+\theta].\end{aligned}\quad(19)$$

In Sec. 3.3 approximations are given to Eqs. (18) and (19) in which only a few terms are kept.

3.2 Discussion of General Formulas

Before discussing approximations to the somewhat cumbersome expressions given above, let us consider the physical implications of these expressions.

In the absence of the external rf field, our system consists of two energy levels having orthogonal wave functions U_a and U_b multiplied respectively by the exponential time-functions $e^{-i\omega_a t}$ and $e^{-i\omega_b t}$. When the field is turned on, the functions multiplying U_a and U_b are no longer simply exponentials, but are in general complicated periodic functions of the time which may be analyzed into a sum of exponential terms multiplied by coefficients, as in Eq. (17)

These terms will be referred to as making up a "spectrum," but not in the usual quantum-mechanical sense of a series of stationary energy states or of transitions between these states. In fact, since the interaction of the molecule with the varying external field introduces a term into the Hamiltonian which explicitly involves the time, there is really no such thing as a stationary state. Rather, there is a mixed (a,b) state whose exact form is a function of the time.

In the following sense, however, it is useful to speak of the spectrum which exists in the presence of the rf field: Assume that a second field of higher frequency is simultaneously applied to the molecule, and that this field is so weak that its effect on the states a and b is negligible compared to that of the rf field. Now assume that a third state, c, exists in the molecule, with the wave function U_c. When the frequency of the second field is resonant with the interval between c and one of the components of the mixed (a,b) state, an occasional transition will be induced which may be observed as the absorption of a photon from the second field. For example, looking at Eq. (16), absorption will occur when its frequency equals $\omega_c+\lambda_a, \omega_c+\lambda_a-2\omega, \cdots\omega_c+\lambda_a-k\omega$. The intensities of these absorptions will be proportional to $|A_0|^2, |A_2|^2\cdots, |A_k|^2$, respectively. It is assumed that the matrix element of the electric dipole

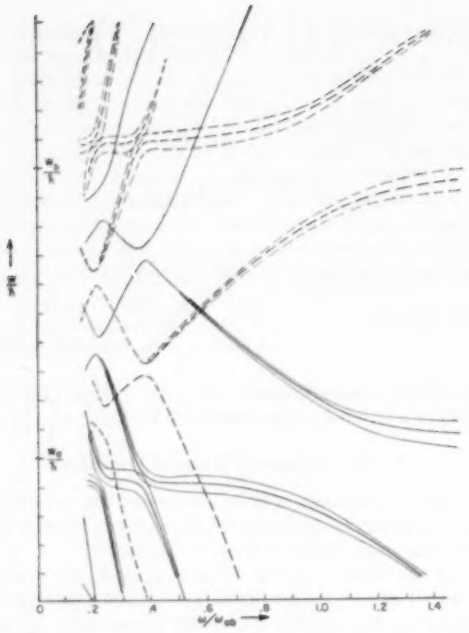


FIG. 2. The frequency spectrum of the components as a function of the applied field frequency, ω, with a field strength such that $\omega_{ab}/\beta_{ab}=5$. The frequency of a component is given by its vertical position while its intensity is roughly represented by the width assigned to it. Very weak components are represented by single lines. Components having wave function U_a are shown by solid lines and components with wave function U_b by dashed lines.

moment between U_a and U_c is not zero. If these absorption lines are observed in a spectrometer, it acts as a "spectrum analyzer" in the usual sense, i.e., an instrument to analyze the frequencies and amplitudes of the Fourier components of a time-varying quantity. This is essentially what was done in the experiment to be described in this paper; the amplitudes of the stronger components being examined critically in the frequency range near resonance.

Figure 2 shows what might be regarded as a typical spectrum. With the field intensity held constant at a moderate value ($\omega_{ab}/\beta_{ab}=5$), the positions of the various components are plotted as functions of ω/ω_{ab}. A rough indication of the relative intensity of a component is given by the thickness of the curve representing it. The solid lines represent a-type components which have the same wave function, U_a, as the unperturbed state "a". The dashed lines represent b-type components. At any particular frequency, pairs of a-type and b-type components alternate. Successive pairs are separated by twice the frequency of the applied field. At certain frequencies the separation between the two components of a pair is a minimum, and somewhere near this frequency the components have equal intensities. However, if the rf field is not too strong one component of each type is usually more intense than all the others and the position of this component changes slowly as the frequency is varied. One may roughly think of this dominant component as being the unperturbed state displaced by an average Stark effect; the other, weaker components being some sort of modulation side bands.

As discussed in the previous section, we have selected the solution of L_a which insures that A_0 will be the largest component as the field strength approaches zero, so the dominant term in the spectrum is $U_a A_0 e^{i\omega_a t}$. If the rf field is turned off, this becomes $U_a A_0 e^{-i\omega_a t}$. Thus the A_0 component undergoes a Stark shift equal to $\lambda_a+\omega_a$ or $\beta_{ab}L_a$. In the presence of an actual field, however, A_0 is not dominant within small frequency ranges near $\omega/\omega_{ab}=1, \frac{1}{3}, \frac{1}{5}\cdots$. For example, if $\omega_{ab}/\beta_{ab}=5$, and $\omega/\omega_{ab}=\frac{1}{3}$, $A_{-3}/A_0=11.05$, and A_{-3} remains greater than A_0 over the range $\frac{1}{3}\leq\omega<0.38$. For weaker fields the frequency at which A_{-3} and A_0 are equal approaches $\omega_{ab}/3$. Similarly, A_{-5} is greater than A_0 for a range of frequencies above $\omega/\omega_{ab}=\frac{1}{5}$. At still lower frequencies several components may be greater than A_0 and the picture mentioned above breaks down altogether as the rf field becomes stronger and its frequency lower.

Our method of choosing L_a at different frequencies does not succeed in making the situation look particularly simple except for very weak fields. However, it is no more complicated than any other method which suggests itself, and it at least avoids some of the ambiguity which would arise from other choices. In Sec. 3.3, we shall obtain approximate solutions by assuming that higher components are negligible. In these solutions L_a is not many-valued and the situation is much simpler. However, for strong rf fields these complications cannot be ignored if accuracy is desired.

Figure 2 is incomplete at the low-frequency end because in this region the continued fractions converge slowly and calculations are very time-consuming. The general behavior in this range is apparent however. As ω decreases, more and more components crowd into a given frequency range on the vertical scale, and the intensity is distributed more uniformly among these components. Although there are more components having appreciable intensity, these components are still restricted to a band of width approximately $2\beta_{ab}$ about the unperturbed frequency of the levels. For several special cases at very low frequencies, it is possible to obtain simple expressions as in Sec. 3.5.

In the resonant region near $\omega/\omega_{ab}=1$, A_0 and A_{-1} are approximately equal and all other components are much smaller. An absorption line involving state "a" therefore appears to split into two nearly intense lines when the resonant rf field is turned on. This is the basis of the "resonant modulation" experiments which are fully discussed in Sec. 4.1.

The appearance of the spectrum in the high-frequency range is quite anomalous. At all the lower frequencies including dc, the most intense a-type and b-type com-

ponents lie farther apart than the unperturbed levels. But Fig. 2 shows that for $\omega \gg \omega_{ab}$ the most intense components are closer together than the unperturbed levels. Thus, the familiar "repulsion of states" which occurs in dc Stark effect becomes an "attraction of states" if a high-frequency field is used. As ω is further increased, the states return to their unperturbed positions and all effects of the field become negligible.

3.3 Approximate Solutions

If the strength of the applied field is not too great, nor its frequency too small, one can obtain useful approximate solutions which are simple functions of the physical parameters by applying Eqs. (10) and (11) and retaining only a small number of quotients in the continued fractions.

First Approximation

The simplest assumption is to approximate Eq. (11) by

$$L_a = \frac{1}{L_a + \dfrac{\omega_{ab}-\omega}{\beta_{ab}}}. \tag{20}$$

This should give a solution which is nearly correct when $\omega \approx \omega_{ab}$, for the first quotient of the first term of Eq. (11) then dominates all of the others. As can be seen in Eq. (10), this amounts to assuming that all the amplitude coefficients except A_0, B_0, A_{-1}, B_1 are negligible.

Equation (20) is quadratic in L_a and its solution is

$$L_a = -\frac{\omega_{ab}-\omega}{2\beta_{ab}} \pm \tfrac{1}{2}\left[\frac{(\omega_{ab}-\omega)^2}{\beta_{ab}^2}+4\right]^{\frac{1}{2}} \tag{21}$$

$$+\text{for } \omega < \omega_{ab}, \quad -\text{for } \omega \geq \omega_{ab}.$$

In accordance with the discussion in Sec. 3.1, the sign is chosen so that L_a remains finite as $\beta_{ab} \to 0$. It follows from (21) that:

$$\lambda_a = -\tfrac{1}{2}(\omega_a+\omega_b) + \tfrac{1}{2}\omega \pm \tfrac{1}{2}[(\omega_{ab}-\omega)^2+4\beta_{ab}^2]^{\frac{1}{2}}. \tag{22}$$

The relative intensities of the components may be obtained from Eqs. (10) and (21).

$$B_{+1}/A_0 = -\beta_{ab}/K_{ab}, \tag{23}$$

where

$$K_{ab} = \tfrac{1}{2}\{(\omega_{ab}-\omega) \pm [(\omega_{ab}-\omega)^2+4\beta_{ab}^2]^{\frac{1}{2}}\}.$$

The sign is chosen according to the same convention as for Eq. (21).

It can be seen from Eqs. (23) and (14) that in this approximation

$$\left|\frac{B_{+1}}{A_0}\right| = \left|\frac{A_{-1}}{B_0}\right| = 1$$

when $\omega = \omega_{ab}$, and this is true for all field strengths.

The expressions for $|T_a|^2$ and $|T_b|^2$ are quite simple in this approximation. Equations (18) and (19) become

$$\begin{aligned} |T_a|^2 &= |A_0|^2 + |B_0|^2(B_{+1}/A_0)^2 \\ &\quad -2|A_0||B_0|(B_{+1}/A_0) \\ &\quad \times \cos\{[(\omega_{ab}-\omega)^2+4\beta_{ab}^2]^{\frac{1}{2}}t+\theta\}, \\ |T_b|^2 &= |A_0|^2(B_{+1}/A_0)^2 + |B_0|^2 \\ &\quad +2|A_0||B_0|(B_{+1}/A_0) \\ &\quad \times \cos\{[(\omega_{ab}-\omega)^2+4\beta_{ab}^2]^{\frac{1}{2}}t+\theta\} \end{aligned} \tag{24}$$

where A_0, B_0, and θ are to be determined by the initial and normalization conditions.

Two particular sets of initial conditions are of most interest. The first is the case where a molecule was definitely known to be in a particular state, say "b," at $t=0$. If we set $|T_a|^2=0$, $|T_b|^2=1$ and $t=0$, (24) gives:

$$\theta = 0, \quad A_0 = \frac{B_{+1}/A_0}{1+(B_{+1}/A_0)^2}, \quad B_0 = \frac{1}{1+(B_{+1}/A_0)^2}. \tag{25}$$

Substituting those values into (24) and making use of (23), we get:

$$|T_a|^2 = \frac{4\beta_{ab}^2}{\gamma^2}\sin^2\frac{\gamma}{2}t, \quad |T_b|^2 = 1-\frac{4\beta_{ab}^2}{\gamma^2}\sin^2\frac{\gamma}{2}t, \tag{26}$$

where

$$\gamma = [(\omega_{ab}-\omega)^2+4\beta_{ab}^2]^{\frac{1}{2}}.$$

Comparison of (26) with (5) shows that they are identical when $g\mu_0 H_\perp = 2\beta_{ab}\hbar$, showing that this approximation is equivalent to assuming a rotating field as Rabi[3] did. This solution is also equal to that of Torrey.[4]

It can be seen from (26) that the molecule oscillates $\gamma/2\pi$ times each second between the two states. When $\omega=\omega_{ab}$ this oscillation frequency is $\beta_{ab}/2\pi$ which is directly proportional to the strength of the applied field. If $\omega \neq \omega_{ab}$ the oscillation frequency is greater, but the probability of finding the system in state "a" never reaches one before it starts to decrease again.

Equation (17) can now be written

$$\begin{aligned} \psi = &\frac{\beta_{ab}}{\gamma}U_a e^{-\frac{1}{2}i(\omega_a+\omega_b)t}\{e^{\frac{1}{2}i(\omega+\gamma)t} - e^{\frac{1}{2}i(\omega-\gamma)t}\} \\ &-\frac{\beta_{ab}}{\gamma}U_b e^{-\frac{1}{2}i(\omega_a+\omega_b)t}\left\{\frac{\beta_{ab}}{K_{ab}}e^{\frac{1}{2}i(-\omega+\gamma)t} + \frac{K_{ab}}{\beta_{ab}}e^{-\frac{1}{2}i(\omega+\gamma)t}\right\}. \end{aligned} \tag{27}$$

Equations (26) and (27) have been derived by assuming a system of molecules known to be in state "b" at $t=0$, a situation which might occur in the case of a molecular beam which has been prepared so that this initial condition holds for all the molecules in the beam.

However, the conditions which must be applied in the "resonant modulation" experiment are somewhat different. Here we have a gas in a wave guide; the

molecules of the gas are colliding with each other and the walls of the wave guide at random and have different histories although they are all influenced by the same rf field. All that we can say about a particular molecule is that at a given time it is equally likely to be either of the two independent states possible in the presence of the rf field. Thus, $|A_0|=|B_0|$ while the phase angle θ is undetermined. If this condition is applied and (24) normalized,

$$|T_a|^2+|T_b|^2=1=2|A_0|^2[1+(B_{+1}/A_0)^2]$$

giving

$$|A_0|=\frac{1}{\sqrt{2}}\frac{1}{[1+(B_{+1}/A_0)^2]^{\frac{1}{2}}}=|B_0|. \tag{28}$$

This, together with (23) gives

$$\begin{aligned}|T_a|^2&=\tfrac{1}{2}\{1-(2\beta_{ab}/\gamma)\cos(\gamma t+\theta)\},\\ |T_b|^2&=\tfrac{1}{2}\{1+(2\beta_{ab}/\gamma)\cos(\gamma t+\theta)\},\end{aligned} \tag{29}$$

where θ is indeterminate for any particular molecule. In considering the entire ensemble of molecules an average must be taken over all values of θ. Equation (29) then shows that at any time there are equal numbers of molecules in the two states.

Equation (17) now becomes

$$\psi=U_a\frac{e^{-\frac{1}{2}i(\omega_a+\omega_b)t}}{\sqrt{2}\gamma K_{ab}}\{K_{ab}e^{\frac{1}{2}i(\omega+\gamma)t}-e^{-i\theta}\beta_{ab}e^{\frac{1}{2}i(\omega-\gamma)t}\}$$

$$-\frac{U_be^{-\frac{1}{2}i(\omega_a+\omega_b)t}}{\sqrt{2}\gamma K_{ab}}\{\beta_{ab}e^{\frac{1}{2}i(-\omega+\gamma)t}+e^{-i\theta}K_{ab}e^{-\frac{1}{2}i(+\omega+\gamma)t}\}. \tag{30}$$

In this form it is easy to determine the spectrum as discussed in Sec. 3.2. If there is a third stationary state of the unperturbed system, having the wave function U_c and for which $|\langle a|\boldsymbol{\mu}\cdot\mathbf{E}|c\rangle|^2$ does not vanish, transitions to this state may be induced by a weak oscillating field of the right frequency. Equation (30) shows that this will occur at two frequencies separated by $\gamma/2\pi$ sec^{-1} and with relative probabilities $|K_{ab}/\beta_{ab}|^2$. Thus, the $a\rightarrow c$ spectrum which is a single line in the unperturbed system splits into two components in this approximation. Absorption lines involving state b will be similarly split. When $\omega=\omega_{ab}$, $K_{ab}=\beta_{ab}=\gamma/2$ so the spectrum of a-type states consists of two equally intense components which are separated in angular frequency by $2\beta_{ab}$ sec^{-1}. The separation of the two components is thus proportional to the intensity of the perturbing field, but their relative amplitudes is independent of it.

When $\omega<\omega_{ab}$, $K_{ab}>\beta_{ab}$ and one component is more intense; however, when $\omega>\omega_{ab}$, $K_{ab}<\beta_{ab}$ and the intensities are interchanged. Thus it appears that ω_{ab} may be determined by varying the frequency of the rf field and observing when the two components are equally intense. In higher approximations this relationship does not hold exactly; the extent of deviation from it depends upon the intensity of the applied field as discussed below.

Second Approximations

The preceding solution is valid only for relatively weak fields with the frequency near resonance. It is simple to find a solution which is valid over a much wider range, including intermediate and high-frequency fields, and which reduces to the previous one near resonance.

Equation (11) converges rapidly when β_{ab}/ω_{ab} and β_{ab}/ω are $\ll 1$. To the first power in these quantities it can be written:

$$L_a=\frac{1}{L_a+\dfrac{\omega_{ab}-\omega}{\beta_{ab}}+\dfrac{\beta_{ab}}{2\omega}}+\frac{\beta_{ab}}{\omega_{ab}+\omega}. \tag{31}$$

Ignoring L_a in some of the denominators is justified if we require a solution for which $L_a\ll\omega/\beta_{ab}$. Equation (31) may be further simplified to

$$L_a^2+L_a\frac{\omega_{ab}-\omega}{\beta_{ab}}-\frac{2\omega_{ab}}{\omega_{ab}-\omega}=0, \tag{32}$$

which has the solutions

$$L_a=-\frac{\omega_{ab}-\omega}{2\beta_{ab}}\pm\frac{1}{2}\left[\frac{(\omega_{ab}-\omega)^2}{\beta_{ab}^2}+4+4\frac{\omega_{ab}-\omega}{\omega_{ab}+\omega}\right]^{\frac{1}{2}}. \tag{33}$$

This differs from Eq. (21) only in the last term inside the brackets, and the same sign conventions are used. Plots of Eq. (33) for several field strengths are given by the dotted curves in Fig. 1. It can be seen that Eq. (33) is a better approximation at the weaker field, but does not show up the multiplicity of solutions or their rapid variations in certain regions.

The amplitude coefficients are:

$$\begin{aligned}\frac{B_{+1}}{A_0}&=-\frac{A_{-1}}{B_0}=\frac{-1}{L_a+\dfrac{\omega_{ab}-\omega}{\beta_{ab}}+\dfrac{\beta_{ab}}{2\omega}},\\ \frac{B_{-1}}{A_0}&=-\frac{A_{+1}}{B_0}=\frac{-\beta_{ab}}{\omega_{ab}+\omega},\\ \frac{A_{+2}}{A_0}&=-\frac{B_{-2}}{B_0}=\frac{\beta_{ab}}{2\omega}\frac{B_{+1}}{A_0},\end{aligned} \tag{34}$$

where L_a is given by Eq. (33).

Very close to resonance when $|\omega_{ab}-\omega|\ll\beta_{ab}$, Eq. (33) approaches (21), the rotating field solution. Even for $\omega=\omega_{ab}$, however, the counter-rotating component makes its presence felt on the amplitude coefficients. Thus, $B_{+1}/A_0\approx 1+(\beta_{ab}/2\omega)$ when $\omega=\omega_{ab}$, showing that the components A_0 and A_{-1} are not quite equal at resonance

in this approximation. Far from resonance where

$$\frac{(\omega_{ab}-\omega)^2}{\beta_{ab}^2} \gg \frac{8\omega_{ab}}{\omega_{ab}+\omega}, \tag{35}$$

Eq. (33) becomes

$$L_a \approx \frac{\beta_{ab}}{\omega_{ab}+\omega} + \frac{\beta_{ab}}{\omega_{ab}-\omega}. \tag{36}$$

The condition of Eq. (35) may be satisfied for fairly strong fields when $\omega > \omega_{ab}$ but only for much weaker fields when $\omega \ll \omega_{ab}$. In addition, the previous condition for this approximation, $\beta_{ab}/\omega \ll 1$, must still be met.

In the intermediate and low-frequency ranges, Eq. (36) or (33) gives a single-valued solution for L_a to replace the complicated one in Fig. 1. This should be adequate except when it is possible to resolve a frequency difference as small as that between A_0 and A_{-3} at their closest approach (Fig. 2).

Better Approximations

To extend the solutions to include terms of order β_{ab}^2/ω^2 and $\beta_{ab}^2/\omega_{ab}^2$ Eq. (11) is approximated by

$$L_a = \cfrac{1}{L_a + \cfrac{\omega_{ab}-\omega}{\beta_{ab}} - \cfrac{1}{L_a - \cfrac{2\omega}{\beta_{ab}}}} + \cfrac{1}{L_a + \cfrac{\omega_{ab}-\omega}{\beta_{ab}}}. \tag{37}$$

All other terms make contributions of third order or higher. It has already been shown that L_a is of zero order near resonance and first order elsewhere. It is only in the resonance region that the additional terms in Eq. (37) give rise to a correction term in the solution for L_a. This solution, obtained in the Appendix 3.3a, is

$$L_a = \rho\left[1 - \frac{\beta_{ab}^2}{4\omega_{ab}^2}\frac{2\rho^2-1}{\rho^2+1}\right], \tag{38}$$

where ρ is the solution for L_a given in Eq. (33). The factor $(2\rho^2-1)/(\rho^2+1)$ varies from $\frac{1}{2}$ to 0 as $|\rho|$ runs from one to $1/\sqrt{2}$. Equation (38) may be used in the region near resonance for which $2\rho^2-1>0$, but outside this range the correction terms is disregarded.

A more accurate solution would retain more of the fractions in Eq. (11) and include terms of higher order in β_{ab}/ω_{ab} and β_{ab}/ω. Third-order terms should show up the rapid variations seen in Fig. 1(b) near $\omega/\omega_{ab} \approx 0.40$; which are associated with the vanishing of $(\omega_{ab}-3\omega)/\beta_{ab}$ in Eq. (11). It would be necessary to retain fifth-order terms to represent in detail the behavior near $\omega/\omega_{ab} = 0.23$.

Since the amount of algebra increases very rapidly with the power of the solution, further calculations have been done numerically. Figures 1 and 2 are based on these results as are the magnitudes of L_a and the amplitude coefficients near resonance. The following procedure was used in calculating each point: A trial value of L_a was used and the two continued fractions in (11) were computed, using two or three denominators in each. If the sum of these two fractions did not give the original trial value of L_a, a new value between the two was tried. This was continued until the trial value of L_a and the computed value agreed to within one digit in the second figure after the decimal point. This procedure was then repeated using several additional denominators. When the result was unaffected by using two additional denominators, the last value calculated was considered final. The subfractions computed in this final calculation were recorded, for they represent the values of $B_{\pm 1}/A_0$, $A_{\pm 2}/B_{\pm 1}$, etc. This can be seen by comparing Eqs. (10) and (11).

3.4a Solution Near Resonance

In this section, we will give a detailed discussion of the solution including terms of second order in β_{ab}/ω_{ab} and β_{ab}/ω. Expressions will be obtained for the relative amplitudes of the components and a condition given for equal intensity of the two main components into which a level splits near resonance.

Making use of Eqs. (10) and (38),

$$\frac{B_{+1}}{A_0} = \cfrac{-\beta_{ab}}{\rho + \cfrac{\omega_{ab}-\omega}{\beta_{ab}} - \rho\cfrac{2\rho^2-1}{\rho^2+1} - \cfrac{1}{\rho - \cfrac{2\omega}{\beta_{ab}}}}, \tag{39}$$

which can be rewritten,

$$\frac{B_{+1}}{A_0} = \cfrac{-\beta_{ab}}{K_{ab} + \cfrac{\beta_{ab}^2}{2\omega} + \cfrac{\beta_{ab}^2}{4\omega_{ab}^2}\cfrac{2-\rho^2}{1+\rho^2}\rho}, \tag{40}$$

where

$$K_{ab} = +(\omega_{ab}-\omega)+\beta_{ab}\rho = \frac{\omega_{ab}-\omega}{2} \pm \frac{1}{2}\left[(\omega_{ab}-\omega)^2 + \frac{8\omega_{ab}\beta_{ab}^2}{\omega_{ab}+\omega}\right]^{\frac{1}{2}}.$$

Use $+$ if $\omega < \omega_{ab}$ and $-$ if $\omega \geqslant \omega_{ab}$. To second order, we may now write

$$\frac{B_{+1}}{A_0} = -\frac{\beta_{ab}}{K_{ab}}\left[1 - \frac{\beta_{ab}}{2\omega}\frac{\beta_{ab}}{K_{ab}} + \frac{\beta_{ab}^2}{4\omega_{ab}^2}\left(\frac{\beta_{ab}^2}{K_{ab}^2} - \frac{2-\rho^2}{1+\rho^2}\frac{\rho}{K_{ab}}\beta_{ab}\right)\right], \tag{41}$$

and

$$\left(\frac{B_{+1}}{A_0}\right)^2 = \frac{\beta_{ab}^2}{K_{ab}^2}\left[1 - \frac{\beta_{ab}}{\omega}\frac{\beta_{ab}}{K_{ab}} + \frac{\beta_{ab}^2}{4\omega_{ab}^2} \times \left(\frac{3\beta_{ab}^2}{K_{ab}^2} - 2\frac{2-\rho^2}{1+\rho^2}\frac{\rho}{K_{ab}}\beta_{ab}\right)\right]. \tag{42}$$

The other components are shown in Appendix 3.4a to be

$$\frac{B_{-1}}{A_0}=-\frac{A_{+1}}{B_0}=-\left[\frac{\beta_{ab}}{2\omega_{ab}}+\frac{\beta_{ab}^2}{4\omega_{ab}^2}\left(\frac{\omega_{ab}-\omega}{\beta_{ab}}-\rho\right)\right],$$

$$\frac{A_{+2}}{A_0}=+\frac{B_{-2}}{B_0}=\frac{B_{+1}}{A_0}\left[\frac{\beta_{ab}}{2\omega_{ab}}+\frac{\beta_{ab}}{4\omega_{ab}^2}\left(2\frac{\omega_{ab}-\omega}{\beta_{ab}}+\rho\right)\right], \tag{43}$$

$$\frac{B_{+2}}{A_0}=-\frac{A_{-2}}{B_0}=-\frac{A_{+2}}{A_0}\frac{\beta_{ab}}{\omega_{ab}-3\omega}=\frac{B_{+1}}{A_0}\frac{\beta_{ab}^2}{4\omega_{ab}^2},$$

$$\frac{A_{-2}}{A_0}=\frac{B_{+2}}{B_0}=+\frac{\beta_{ab}^2}{4\omega_{ab}^2}.$$

We define the "splitting," γ, as the angular frequency separating the components A_0 and A_{-1} or B_0 and B_{+1} near resonance when they have approximately equal intensities. Equation (17) shows that A_0 and A_{-1}, the two predominant a-type components, have relative amplitudes B_{+1}/A_0 and are separated by

$$\gamma=2\beta_{ab}L_a+\omega_{ab}-\omega. \tag{44}$$

Now make use of Eqs. (38) and (33)

$$\gamma=2\beta_{ab}\rho\left[1-\frac{\beta_{ab}^2}{4\omega_{ab}^2}\frac{2\rho^2-1}{\rho^2+1}\right]+\omega_{ab}-\omega,$$

$$|\gamma|=\beta_{ab}\left\{\left[\frac{(\omega_{ab}-\omega)^2}{\beta_{ab}^2}+\frac{8\omega_{ab}}{\omega_{ab}+\omega}\right]^{\frac{1}{2}}-\frac{\beta_{ab}^2}{2\omega_{ab}^2}\frac{2\rho^2-1}{\rho^2+1}\right\}. \tag{45}$$

At $\omega=\omega_{ab}$

$$|\gamma|=2|\beta_{ab}|\left[1-\frac{\beta_{ab}^2}{8\omega_{ab}^2}\right].$$

Equation (24) shows that the amplitude with which the molecule oscillates between states a and b is proportional to $A_0B_0(B_{+1}/A_0)$. In the case where a molecule is initially known to be in one of the states this amplitude is, from Eq. (25), $(B_{+1}/A_0)^2[1+(B_{+1}/A_0)^2]^{-1}$ which has a maximum when $B_{+1}/A_0=1$. The frequency at which this occurs may be obtained by setting $B_{+1}/A_0=1$ in Eq. (50) of Sec. 4.2. Solving this equation gives the condition for resonance:

$$\frac{\omega}{\omega_{ab}}=1+\frac{\beta_{ab}^2}{\omega_{ab}^2}, \tag{46}$$

where $\hbar\omega_{ab}$ is the energy difference between the two levels. As was pointed out in Sec. 2.4, this result has been obtained by Bloch and Siegert.[5] For stronger fields, the shift in resonance may be calculated numerically from Eqs. (10) and (11) by determining at each field strength the frequency at which B_{+1}/A_0 equals one.

3.5 Low-Frequency Perturbation

If the frequency of the rf field is low enough, the problem may be treated quasistatically as outlined in Sec. 2.6. We will first consider the part of the solution which reduces to state "a" at zero rf field intensity.

For a dc electric field, the solution is

$$\psi=N\left[U_a-\frac{4\beta_{ab}}{+(\omega_{ab}/2)+\frac{1}{2}(\omega_{ab}^2+16\beta_{ab}^2)^{\frac{1}{2}}}U_b\right]$$

$$\times\exp\left\{i\left[\frac{\omega_a+\omega_b}{2}-\tfrac{1}{2}(\omega_{ab}^2+16\beta_{ab}^2)^{\frac{1}{2}}\right]t\right\}, \tag{47}$$

where N is a normalization factor. If the frequency is small and $\beta_{ab}\ll\omega_{ab}$, we may replace β_{ab} by $\beta_{ab}\cos\omega t$. Then

$$U_a\sum A_k e^{i(\lambda_a-k\omega)t}$$

$$\approx NU_a e^{i\omega_a t}\exp\left(-i\int_0^t\frac{4\beta_{ab}^2\cos^2\omega t}{\omega_{ab}}dt\right)$$

$$\approx NU_a\exp\left[i\left(\omega_a-\frac{2\beta_{ab}^2}{\omega_{ab}}\right)t\right]\exp\left[-i\frac{\beta_{ab}^2}{\omega_{ab}}\frac{\sin2\omega t}{\omega}\right].$$

Clearly $\lambda_a=\omega_a-(2\beta_{ab}^2/\omega_{ab})$, and

$$A_n=\frac{1}{2\pi}N\int_{-\pi}^{\pi}\exp\left[i\left(2n\omega t-\frac{\beta_{ab}^2}{\omega_{ab}}\frac{\sin2\omega t}{\omega}\right)\right]d(2\omega t),$$

where k is any even integer and n any integer. The integral is familiar, and A_n may be written

$$A_n=NJ_n\left(\frac{\beta_{ab}^2}{\omega_{ab}\omega}\right).$$

Thus the U_a wave function has a spectrum given by

$$\sum_{n=-\infty}^{+\infty}J_n\left(\frac{\beta_{ab}^2}{\omega_{ab}\omega}\right)e^{-2i n\omega t}\exp\left[i\left(\omega_a-\frac{2\beta_{ab}^2}{\omega_{ab}}\right)t\right]$$

showing the frequency-shift of the main component and the side bands, whose intensities fall off as

$$|J_n(\beta_{ab}^2/\omega\omega_{ab})|^2.$$

This is the solution for second-order Stark effect described and observed by Townes and Merritt.[8]

For first-order Stark effect, by setting $\omega_{ab}=0$ in Eq. (47) and proceeding as before one can show that the main component is unshifted while the side bands are separated by the ω instead of 2ω and fall off as $|J_n(2\beta_{ab}/\omega)|^2$. When ω_{ab} and β_{ab} are of comparable magnitude, there are no such simple expressions for the positions and intensities of the components.

IV. RESONANT MODULATION

4.1 Discussion of Resonant Modulation Experiment in OCS

Measurements were carried out near resonance to check the theory and also to determine the practicality of using resonant modulation as a precision method of determining the energy difference between certain molecular levels. For instance, one might wish to determine the interval between two hyperfine levels corresponding to the same rotational angular momentum by observing a microwave transition to another rotational state while inducing transitions between the hyperfine levels by an rf field.

As a first test, however, it was decided to use the *l*-type doublets of the linear molecule OCS (carbonyl sulfide). These levels satisfy our assumption of isolated pairs of states and also provide fairly strong microwave absorption lines to observe. They are also uncomplicated by nuclear interactions.

Figure 3(a) shows the levels of the OCS molecule which are of interest. The measured microwave frequencies of the two allowed transitions $a \rightarrow c$ and $b \rightarrow d$ are indicated on the diagram. The energy intervals (a, b) and (c, d) are much smaller; they arise from the vibration-rotation interaction in which coriolis forces remove the degeneracy between the two orthogonal bending vibrations of the linear OCS molecule.[11] These interactions increase with rotational angular momentum, J, as $J(J+1)$ so the (c, d) separation should be three times as great as for (a, b). On the basis of this assumption, and previous measurements[12] of the frequencies $a \rightarrow c$ and $b \rightarrow d$, the (a,b) and (c,d) separations are determined to be 12.76 Mcps and 38.28 Mcps respectively.

The resonant modulation experiment is carried out by observing the two absorption lines $a \rightarrow c$ and $b \rightarrow d$ in a conventional Stark-type[13] microwave spectroscope while the rf field is applied to the molecules. A schematic diagram of the apparatus is shown in Fig. 5. With the OCS gas at a pressure of about 10^{-2} mm in the wave guide of the spectroscope, an rf field with an amplitude of several volts is applied to the same septum as the 100 kc square-wave voltage used in spectrometers of this type. This subjects the gas to a fairly uniform rf field of the order of 10 volts-cm intensity. The 100 kc square-wave Stark voltage simultaneously applied to the septum plays no part in the resonant modulation process, but is merely part of the microwave detection system.

When the frequency, ν, of the rf voltage is near either ν_{ab} or ν_{cd}, each absorption line appears to split into two components. The relative intensity of these components is sensitive to the frequency of the rf field, and

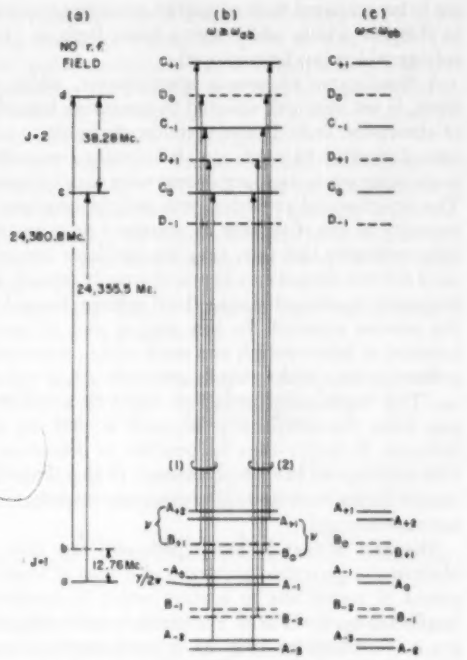


FIG. 3. (a) The unperturbed OCS energy levels of interest in the resonance modulation experiment. (b) The components are shown when the applied field frequency is slightly above resonance. The transition within groups (1) and (2) occur at the same microwave frequency. (c) The components are shown relabeled according to our convention when $\omega > \omega_{ab}$.

if the strength of the field is held constant while its frequency is varied through resonance, first one and then the other component becomes more intense. This is illustrated in Fig. 2.

In the first approximation, valid for weak rf fields, the relative intensity of the two components is independent of the amplitude of the field, the two components being equally intense when ν is exactly equal to ν_{ab} or ν_{cd}. This suggests that one could measure ν_{ab} directly by determining the rf frequency at which the two components are equally intense. However, the higher approximations in Sec. 3.3 predict, and experimental results confirm, that for greater rf field strength the relative intensity of the components depends upon field strength as well as frequency. One would like to use as weak a field as possible in order to minimize the dependence on field intensity and also because, as discussed in Sec. 4.3, the resonance is sharper for weak fields and one could expect to measure ν_{ab} more accurately. However, as the field strength is decreased, the separation of the components also decreases. Although it has been assumed in the theory that the lines are perfectly sharp, they, of course, actually have a finite width, in our case about $\frac{1}{4}$ megacycle, due mainly to pressure broadening. If the intensities of the components

[11] H. H. Nielsen, Revs. Modern Phys. **23**, 90 (1951).
[12] P. Kisliuk and C. H. Townes, Table of Microwave Lines, National Bureau of Standards Circular 518, 1952 (unpublished).
[13] R. H. Hughes and E. B. Wilson, Phys. Rev. **71**, 562 (1947).

are to be compared their separation must be comparable to this line width, which sets a lower limit on the rf voltage which can be employed.

A Stark-type microwave spectrometer, while sensitive, is not very well adapted to measuring intensities of absorption lines. It was therefore desirable to minimize difficulties by confining all intensity comparisons to deciding when two components were equally intense. The experimental procedure was to hold constant the intensity of the rf voltage as measured on a vacuum-tube voltmeter and then vary the oscillator frequency until the two components appeared equally intense. This frequency was recorded as ν_0, the rf voltage changed and the process repeated. In this way, a plot of ν_0 as a function of field strength was made which, when extrapolated to zero field strength, gives the actual value of ν_{ab}. This complicated procedure might be simplified if one knew the relationship between ν_0 and the field strength. It might then be possible to determine ν_{ab} with only one or two measurements. In this connection two problems need to be discussed, one theoretical and one experimental.

The first is that in our experiment, one does not observe the perturbed spectrum of a pair of levels by means of transitions to a state which is completely unaffected by the rf field. For example, when observing the $a \rightarrow c$ transition (Fig. 3) in the presence of an rf field which is resonant with ν_{ab}, the field affects states "c" and "d" as well as "a" and "b." Although the effect of the rf field on "c" will not be so great as on "a" because ν is not resonant with ν_{cd}, it may still be great enough to change the results and must be calculated. This is done in the next section.

The experimental difficulty arose from the fact that we were unable to make a reliable measurement of the amplitude of the rf voltage. Equation (45) relates the separation of the two components, or "splitting" to the amplitude of the applied voltage, assuming that one knows the matrix element between states a and b. The matrix elements in this case can be calculated[11] and an experimental check of the above relation gave a discrepancy of the order of 30%. It was assumed that this discrepancy was due to the vacuum-tube voltmeter giving an incorrect reading of the actual effective voltage across the wave guide. Smaller inconsistencies between separate measurements of the voltage also indicated that they were unreliable. As the basic purpose of this experiment was to study the behavior of "resonant modulation," rather than to measure frequency intervals, it was decided that the "splitting" itself would be used as means of determining the voltage. This "splitting" then was directly measured by using a microwave frequency standard. This is a rather inconvenient procedure which in effect nullifies one of the advantages of using resonant modulation; namely, the ability to make frequency measurements in the rf rather than the microwave range. It seems likely that a way to make reliable voltage measurements could be found and the need for a microwave frequency standard eliminated.

4.2 Theory of Resonant Modulation in OCS

Figure 3(b) shows the spectrum of the l-type doublets of OCS when subjected to an rf field resonant with ν_{ab}. The spectrum of the (a, b) state is characteristic of the resonant condition. The (c, d) spectrum is nonresonant since $\nu \approx \nu_{cd}/3$. All the A components have the same wave function, $U_a(x,y,z)$, as the unperturbed "a" state. Similarly, the B, C, and D components have the wave functions of the other unperturbed states.

The components A_0 and A_{-1}, B_0, and B_{+1} are about equally intense and occur in pairs which are "split" by a frequency difference which is very nearly proportional to the rf field intensity. The magnitude of this "splitting" is given in Eq. (45). The weaker components such as A_{+2} and A_{+1}, B_{-2}, and B_{-1} also occur in pairs of approximately equal intensity. In the upper state the intense C_0 and D_0 components are shifted relatively little from their unperturbed positions, while the other components are weaker by a factor of the order β_{cd}/ω_{cd}.

For the l-type doublets, the matrix elements of the dipole moment vanish, except between the states a and c or b and d. Therefore, as shown in Fig. 3(a), only two microwave absorption lines are observed when there is no rf field. In the presence of the field, corresponding to each of these lines, absorption occurs at two different microwave frequencies which are separated by the "splitting" frequency, $\gamma/2\pi$. There are many pairs of components separated by each of these frequencies, but we will consider only the five pairs for which the microwave absorption varies as the second power or lower of the rf field. The problem is to calculate the precise frequency, ν_0 at which transitions (1) and (2) are equally intense.

In absorption spectroscopy, the radiation used is usually so weak that it induces transitions in the absorbing atoms or molecules of a gas at a much lower rate than collisions occur between the molecules. As discussed in Sec. 2.2, the radiation then produces very little change in the relative populations of the molecular states. Under these conditions the fraction of the radiation absorbed at the peak of an absorption line is independent of the intensity of the radiation and is proportional to $|\langle\psi_a|\mu|\psi_c\rangle|^2$ where $\langle\psi_a|\mu|\psi_c\rangle$ is the matrix element of the electric or magnetic dipole moment between the initial and final states. For example, in Fig. 3(a) microwave absorption lines are shown for the transitions from state a to c and from b to d, when there is no rf field. The electric dipole matrix elements and hence intensities of these lines are equal.

In the presence of the rf field, the $J=1$ state has the wave function

$$\Psi_{ab} = U_a T_a(t) + U_b T_b(t),$$

where $T_a(t)$ and $T_b(t)$ may be analyzed into Fourier components as in Eq. (16). Similarly the $J=2$ state

has the wave function

$$\Psi_{cd}=U_cT_c(t)+U_dT_d(t).$$

Figure 3(b) shows the "spectrum" corresponding to those two complex states when ω is slightly greater than ω_{ab}. Each horizontal line represents a Fourier component of T_a, T_b, T_c, or T_d; its frequency is indicated by the vertical position of the line.

When the rf field is on, the probability of a $J=1\rightarrow 2$ transition is still proportional to the square of the dipole matrix element between the initial and final states, although these are no longer stationary states. Thus the absorption is proportional to

$$|\langle\Psi_{ab}|\boldsymbol{\mu}\cdot\mathbf{E}_me^{i\omega mt}|\Psi_{cd}\rangle|^2,$$

where $\mathbf{E}_me^{i\omega mt}$ is the microwave field. In Fig. 3(b), we see that components B_{+1} and D_0 are separated by the same frequency as A_0 and C_{-1}, B_{-1} and D_{+1}, etc. If the microwave frequency is adjusted to resonate with this frequency, absorption occurs which is proportional to

$$|(B_{+1}D_0+B_{-1}D_{+1})\beta_{bd}+[A_0C_{-1}+B_{-1}D_{+1}+A_{-2}C_0)\beta_{ac}|^2,$$

Other pairs of components which are separated by the same frequency are disregarded here because their contribution is small. The intensity of transition (2), which occurs at a frequency $\gamma/2\pi$ greater, is given by a similar expression. For this set of levels $\beta_{bd}=\beta_{ac}$, which may be derived from the theory of the l-type doublets but is also demonstrated by the fact that the two microwave lines are observed to be equally intense when there is no rf field.

The condition that the two absorption lines (1) and (2) be equally intense may be written

$$\begin{aligned}&A_{+2}C_{+1}^*+B_{+1}D_0+A_0C_{-1}+B_{-1}D_{+1}+A_{-2}C_0\\&\quad=A_{+1}C_{+1}+B_0D_0+A_{-1}C_{-1}+B_{-2}D_{+1}+A_{-3}C_0.\end{aligned}\tag{48}$$

Divide Eq. (46) by D_0 and make use of the relations between coefficients in Eq. (15). Also set $A_0=B_0$ and $C_0=D_0$. Then

$$\begin{aligned}&\frac{B_{+1}}{A_0}-\frac{D_{+1}}{C_0}+\frac{B_{-1}}{A_0}\frac{D_{+1}}{C_0}+\frac{A_{-2}}{A_0}-\frac{A_{+2}}{A_0}\frac{D_{-1}}{C_0}\\&\quad=1-\frac{B_{+1}}{A_0}\frac{D_{+1}}{C_0}\frac{A_{+2}}{A_0}\frac{D_{+1}}{C_0}-\frac{B_{+2}}{A_0}\frac{B_{-1}}{A_0}\frac{D_{-1}}{C_0}.\end{aligned}\tag{49}$$

Now, making use of such identities as

$$B_{+2}/A_0=(B_{+2}/A_{+2})(A_{+2}/B_{+1})(B_{+1}/A_0),$$

we obtain

$$\begin{aligned}&\frac{B_{+1}}{A_0}\left[1-\frac{D_{+1}}{C_0}+\frac{B_{+2}}{A_{+2}}\frac{A_{+2}}{B_{+1}}-\frac{A_{+2}}{B_{+1}}\frac{D_{+1}}{C_0}-\frac{A_{+2}}{B_{+1}}\frac{D_{-1}}{C_0}\right]\\&\quad=1-\frac{D_{+1}}{C_0}-\frac{A_{-2}}{B_{-1}}\frac{B_{-1}}{A_0}-\frac{B_{-1}}{A_0}\frac{D_{+1}}{C_0}+\frac{B_{-1}}{A_0}\frac{D_{-1}}{C_0}.\end{aligned}\tag{50}$$

The formulas for the A and B coefficients, which apply in the resonance region, are given in Eqs. (41)–(44). The C and D coefficients may be obtained from Eqs. (34) by replacing a by c and A by C; also b by d and B by D. For example, corresponding to $B_{-1}/A_0=-\beta_{ab}/\omega_{ab}+\omega$ one obtains $D_{-1}/C_0=-\beta_{cd}/\omega_{cd}+\omega$.

We will substitute the appropriate expressions for the coefficients into (49) and solve for the frequency, ω_0, at which transitions (1) and (2) are equally intense. In Eq. (50), B_{+1}/A_0 is approximately equal to one; but the other coefficients are smaller, being of the order β_{ab}/ω_{ab} or β_{cd}/ω_{cd}.

In Appendix 4.2a, it is shown that near resonance Eq. (41) can be simplified to

$$\frac{B_{+1}}{A_0}=1+\frac{\beta_{ab}}{2\omega_{ab}}+\frac{\beta_{ab}^2}{8\omega_{ab}^2}+\frac{\omega_{ab}-\omega}{\beta_{ab}}\left(\frac{1}{2}+\frac{\beta_{ab}}{4\omega_{ab}}\right)+\frac{1}{8}\frac{(\omega_{ab}-\omega)^2}{\beta_{ab}^2},\tag{51}$$

where terms up to the second power of the field strength have been kept. Equation (50) is then solved, giving

$$\omega_0=\omega_{ab}+\frac{5}{3}\frac{\beta_{ab}^2}{\omega_{ab}}+\frac{4\beta_{ab}\beta_{cd}}{\omega_{cd}-\omega_{ab}}+\frac{\beta_{ab}^3}{\omega_{ab}^2}+\frac{2\beta_{ab}^2\beta_{cd}}{\omega_{ab}(\omega_{cd}-\omega_{ab})}.\tag{52}$$

For the l-type doublets[11] in OCS, $\beta_{cd}=\beta_{ab}/3$ and $\omega_{cd}=3\omega_{ab}$. Equation (52) can then be written

$$\nu_0=\nu_{ab}+\frac{5}{3\nu_{ab}}\left(\frac{\beta_{ab}}{2\pi}\right)^2+\frac{4}{3\nu_{ab}^2}\left(\frac{\beta_{ab}}{2\pi}\right)^3.\tag{53}$$

The measured quantities in our experiment were ν_0 and the "splitting," γ. By solving Eq. (45) for β_{ab} as an explicit function of γ, one obtains

$$\frac{\beta_{ab}}{2\pi}=\frac{\gamma/2\pi}{2\left[1+\dfrac{(2\pi)^2(\nu_{ab}-\nu)^2}{\gamma^2}+\dfrac{\nu_{ab}-\nu}{2\nu_{ab}}-\dfrac{\gamma^2}{64\pi^2\nu_{ab}^2}\right]^{\frac{1}{2}}}.\tag{54}$$

Figure 4 is a plot of the measured values of ν_0 against $\beta_{ab}/2\pi$. The values of β_{ab} were obtained by measuring the "splitting" with the aid of a microwave frequency standard and then using Eq. (54). The theoretically predicted curve for ν_0 as a function of β_{ab} is also shown in Fig. 4.

If ν_{ab} is set equal to 12.78 Mc, Eq. (53) becomes

$$\nu_0=12.78+0.131\left(\frac{\beta_{ab}}{2\pi}\right)^2+0.0082\left(\frac{\beta_{ab}}{2\pi}\right)^3.\tag{55}$$

4.3 Stark Effect Above Resonance

In Sec. 3.2, it was mentioned that when $\omega>\omega_{ab}$ the separation of the states appears to be less than in the absence of the field; this can now be proved. Equation (36) shows that L_a is negative above resonance and it follows from the definition of L_a that the A_0 and B_0 components are separated by less than ω_{ab}. In addition,

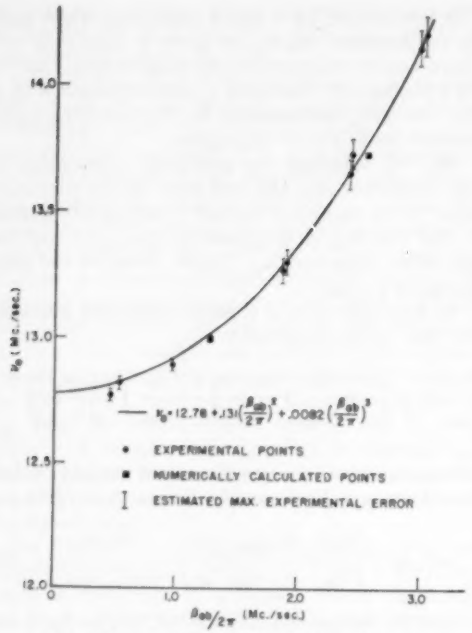


FIG. 4. Measured and calculated values of ν_0 as a function of the field intensity.

by Eq. (34), A_{-1} and B_{+1} are seen to become small when $\omega \gg \omega_{ab}$, showing that A_0 and B_0 are the dominant components. This anomalous behavior has been experimentally observed. Although the microwave absorption lines in Fig. 3(a) move closer together if a constant electric field is turned on, they move apart if an rf field of about 16 megacycles is applied.

4.4 Discussion of Experimental Results

In Fig. 4, the theoretical and experimental results are compared. The solid curve is a plot of Eq. (55) while the circled points were obtained experimentally. ν_{ab} was set equal to 12.78 Mc in Eq. (55) because this is the value calculated from the frequencies of the microwave lines[12] when it is assumed in accordance with the theory of l-type doubling[10] that $\omega_{cd}=3\omega_{ab}$. The theory and experiment are seen to agree within the experimental uncertainty. The data seems to indicate that $\nu_{ab}=12.76$ Mc might be a better fit than $\nu_{ab}=12.78$ Mc, but this difference is not experimentally significant here.

The estimated maximum experimental error which is indicated by the vertical lines on the figure increases with the field strength for several reasons. One is that the components are separated by a greater frequency when the field is stronger; the greater the separation between the components the more likely it is that spurious intensity differences will be introduced by reflections in the waveguide or by variations of the power output of the klystron with frequency. In addition the intensities of the components vary more rapidly with rf frequency when the field strength is small. A given uncertainty in determining the intensities of the components thus results in a smaller frequency uncertainty for a weak rf field than for a strong one. Let R be the minimum detectable fractional difference in intensity of the two components. It follows from Eq. (30) that

$$|\omega_{ab}-\omega_0| = (R/2)\gamma.$$

For a given R, the error in $|\omega_{ab}-\omega_0|$ is thus proportional to the "splitting," γ. For example, if the components can be made equal to within 4% $R=0.04$ and if the "splitting" is one Mc, the uncertainty in ω_0 is about 0.02 Mc. This is approximately the uncertainty is measuring microwave frequencies when the line width is of the order of $\frac{1}{2}$ Mc. If the line width could be reduced, a weaker rf field could be used, thus reducing the "splitting" and increasing the resolution.

To summarize: resonant modulation provides a method of measuring the energy difference between two energy levels. The precision is approximately equal to that with which the microwave absorption line can be located. However, if it is possible to measure accurately the voltage applied to the wave guide, the energy separation between closely spaced levels can be obtained by making frequency measurements in the rf rather than the microwave region. In addition, under some conditions data could be obtained by resonant modulation which could not be obtained by other techniques. For example, in Fig. 3(a), if there were no observable microwave transition from b to d, or the relationship between ω_{ab} and ω_{cd} were not known, one could not measure ω_{ab}. Using resonant modulation, however, it is possible to determine ω_{ab} by observing a single microwave absorption line.

4.5 Details of the Experimental Apparatus and Procedure

Figure 5 is a schematic diagram of the experimental arrangement used. The apparatus outside the dashed area is a conventional Stark-modulation microwave spectroscope. The apparatus inside the area was added for this experiment. It consists simply of an oscillator, a vacuum-tube voltmeter, and a circuit for tuning out the capacitance of the wave guide.

The oscillator is a standard Hartley circuit using an 829 tube and having a number of plug-in coils so it can oscillate in the 11–14 Mc region near ν_{ab} and also the 35-45 Mc region near ν_{cd}.

The vacuum-tube voltmeter used was a General Radio type 1800A which is claimed to be accurate for frequencies much higher than any used in this experiment.

The tuning circuit was used to increase the impedance presented to the oscillator by the capacitance between the wave guide and septum. This is about 1000 $\mu\mu$f which represents a reactance of about 12 ohms at 13 Mc and 4 ohms at 39 Mc. If 10 volts were maintained

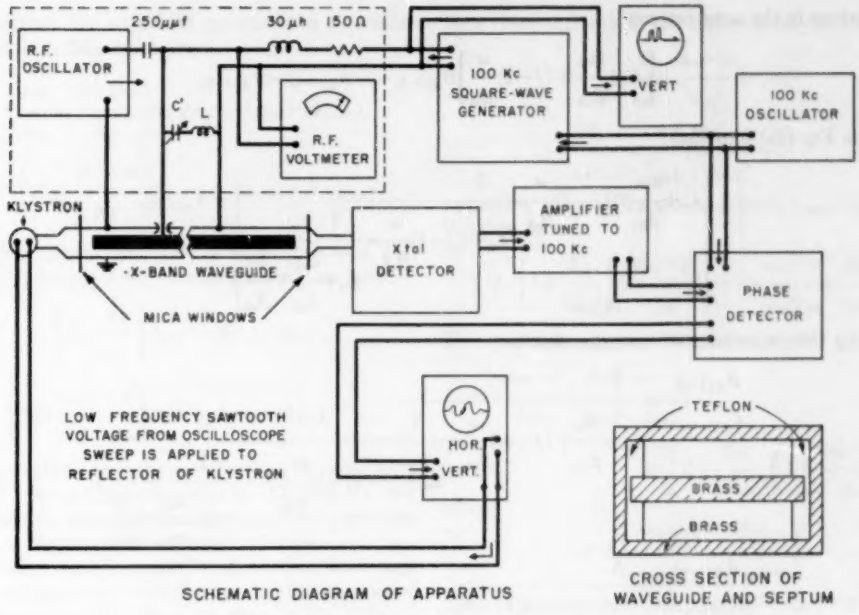


FIG. 5. Schematic diagram of experimental apparatus.

across the septum, rf currents of about 0.83 and 2.50 amperes would be drawn, requiring a sizable oscillator. Instead, in the 13 Mc region this wave-guide capacitance, C, is resonated with a circuit consisting of an inductance and a small variable capacitance in series. These are L and C' in Fig. 5.

The effective capacitance across the coil is the series combination of C and C' which is about equal to C' if $C' \ll C$. The product LC' is adjusted to resonate with the oscillator frequency, thus increasing the impedance of the oscillator's load by a factor equal to the Q of the resonant circuit.

The wave guide is about 1 meter long and is fed near its center. Its effective length of 0.5 meter is thus considerably less than a quarter-wavelength even at 39 Mc so the rf field strength is fairly uniform along its length. The design of the septum, as shown in Fig. 5, is such that the field is also quite uniform over the cross section of the guide.

The 250 $\mu\mu$f condenser isolates the rf oscillator at low frequencies and prevents the output coil of the oscillator from short-circuiting the square-wave generator. Similarly, the 30 μh inductance keeps the square-wave generator from loading the circuit at 13 or 39 Mc. The 150 ohm resistor critically damps the resonant circuit formed by this inductance and the wave-guide capacitance.

The microwave frequency standard used has been described before.[14]

[14] See for example: S. Geschwind, Ann. N. Y. Acad. Sci. **55**, 751 (1952).

ACKNOWLEDGMENTS

The authors are very grateful for help from Mr. George C. Dousmanis, who carried out the numerical calculations reported here, and from Miss Carlotta Heller who assisted with preparation of the manuscript. We would also like to thank the many members of the Columbia Radiation Laboratory who contributed to the construction and maintenance of the experimental apparatus.

APPENDIX 3.1a. PROOF OF EQS. (10)

In Eq. (8) assume that if k and l are greater than some integer j, A_k and B_l are sufficiently small to be neglected. Assuming j is odd and positive

$$\left[L_a+\frac{\omega_{ab}}{\beta_{ab}}\mp j\frac{\omega}{\beta_{ab}}\right]B_{\pm j}=-A_{\pm(j-1)},$$

$$\left[L_a\mp(j-1)\frac{\omega}{\beta_{ab}}\right]A_{\pm(j-1)}=B_{\pm(j-2)}-B_{\pm j}$$

$$=-B_{\pm(j-2)}+\frac{A_{\pm(j-1)}}{L_a+\dfrac{\omega_{ab}}{\beta_{ab}}\mp j\dfrac{\omega}{\beta_{ab}}},$$

$$\left[L_a\mp\frac{(j-1)\omega}{\beta_{ab}}-\frac{1}{L_a+\dfrac{\omega_{ab}}{\beta_{ab}}\mp j\dfrac{\omega}{\beta_{ab}}}\right]A_{\pm(j-1)}=B_{\pm(j-2)}. \quad (2a)$$

Now continue in the same manner

$$\left[L_a+\frac{\omega_{ab}}{\beta_{ab}}\mp(j-2)\frac{\omega}{\beta_{ab}}\right]B_{\pm(j-2)}=-A_{\pm(j-2)}-A_{\pm(j-1)}. \tag{3a}$$

Substitute Eq., (2a) into (3a)

$$\left[L_a+\frac{\omega_{ab}}{\beta_{ab}}\mp(j-2)\frac{\omega}{\beta_{ab}}-\cfrac{1}{L_a\mp(j-1)\dfrac{\omega}{\beta_{ab}}-\cfrac{1}{L_a+\dfrac{\omega_{ab}}{\beta_{ab}}\mp\dfrac{j\omega}{\beta_{ab}}}}\right]=A_{\pm(j-3)}.$$

Continuing this procedure, we can generalize to

$$\frac{B_{\pm(j-m)}}{A_{\pm(j-m-1)}}=\cfrac{1}{L_a+\dfrac{\omega_{ab}}{\beta_{ab}}\mp(j-m)\dfrac{\omega}{\beta_{ab}}-\cfrac{1}{L_a\mp(j-m+1)\dfrac{\omega}{\beta_{ab}}-\cdots\cfrac{-1}{L_a+\dfrac{\omega_{ab}}{\beta_{ab}}\mp\dfrac{j\omega}{\beta_{ab}}}}},$$

$$\frac{A_{\pm(j-n)}}{B_{\pm(j-n-1)}}=\cfrac{1}{L_a\mp(j-n)\dfrac{\omega}{\beta_{ab}}-\cfrac{1}{L_a+\dfrac{\omega_{ab}}{\beta_{ab}}\mp(j-n+1)\dfrac{\omega}{\beta_{ab}}-\cdots-\cfrac{1}{L_a\mp\dfrac{j\omega}{\beta_{ab}}}}},$$

where m is any positive even integer and n any positive odd integer. Now let $j-m=l$ and $j-n=k$, and let $j\to\infty$ while holding k and l constant. Then the above approaches Eq. (10).

APPENDIX 3.1b. EVALUATION OF EQ. (11) WHEN $\omega=0$

When $\omega=0$ Eq. (11) becomes

$$\frac{L_a}{2}=\cfrac{1}{L_a+\dfrac{\omega_{ab}}{\beta_{ab}}-\cfrac{1}{L_a-\cfrac{1}{L_a+\dfrac{\omega_{ab}}{\beta_{ab}}-\cdots}}}.$$

Since this continued fraction repeats indefinitely, eliminating a finite number of terms does not change its value, and the above equation can be written

$$\frac{L_a}{2}=\cfrac{1}{L_a+\dfrac{\omega_{ab}}{\beta_{ab}}-\cfrac{1}{L_a-\dfrac{L_a}{2}}}.$$

This is equivalent to the quadratic equation

$$L_a^2+\frac{\omega_{ab}}{\beta_{ab}}L_a+4=0,$$

whose solution is

$$L_a=-\frac{\omega_{ab}}{2\beta_{ab}}\pm\left[\left(\frac{\omega_{ab}}{2\beta_{ab}}\right)^2+4\right]^{\frac{1}{2}}.$$

This is the same result which can be obtained by direct diagonalization of the time-independent Schrödinger equation.

APPENDIX 3.1c. DERIVATION OF OTHER SOLUTIONS OF EQ. (8)

Assume that L_a' and B_l', $\cdots A_k'\cdots$ satisfy Eq. (8) and that they are related to another set of quantities L_a, $B_l\cdots$, $A_k\cdots$ by

$$L_a'=L_a-\frac{\omega_{ab}-\omega}{\beta_{ab}},\quad B_l'=A_{1-l},\quad A_k'=-B_{1-k}. \tag{12}$$

Since the primed quantities are assumed to satisfy (8)

$$\left[L_a'-\frac{k\omega}{\beta_{ab}}\right]A_k'=-B_{k-1}'-B_{k+1}',$$

$$\left[L_a'+\frac{\omega_{ab}}{\beta_{ab}}-l\omega\right]B_l'=-A_{l+1}'-A_{l+1}'.$$

Now substitute the unprimed quantities to determine what equations they must obey.

$$\left[L_a+\frac{\omega_{ab}}{\beta_{ab}}+\frac{(k-1)\omega}{\beta_{ab}}\right]B_{1-k}=-A_{1-k+1}-A_{1-k-1},$$

$$\left[K_a+\frac{(l-1)\omega}{\beta_{ab}}\right]A_{1-l}=-B_{1-l+1}-B_{1-l-1}.$$

Taking the complex conjugate of these equations, and rewriting by letting $l'=1-k$ and $k'=1-l$, we have:

$$\left[L_a-\frac{k'\omega}{\beta_{ab}}\right]A_{k'}=-B_{k'-1}-B_{k'+1},$$

$$\left[L_a+\frac{\omega_{ab}}{\beta_{ab}}-\frac{l'\omega}{\beta_{ab}}\right]B_{l'}=-A_{l'-1}-A_{l'+1}.$$

These equations have the same form as (8) so we have shown that the primed quantities in (12) define a new solution of (8), provided that the unprimed quantities from which they are derived form a solution.

In a similar manner, it can be shown that from each of these solutions can be derived an infinite number of other solutions, of (8), which are given in Eq. (14).

Deriving Solutions of Eq. (9) from those of (8)

Rewrite (9),

$$\left[L_b-\frac{\omega_{ab}}{\beta_{ab}}-\frac{l\omega}{\beta_{ab}}\right]=-\frac{B_{l-1}}{B_0}-\frac{B_{l+1}}{B_0},$$

$$\left[L_b-\frac{k\omega}{\beta_{ab}}\right]=-\frac{A_{k-1}}{B_0}-\frac{A_{k+1}}{B_0}.$$

Now assume L_b, $A_l/A_0\cdots$, $B_k/B_0\cdots$, satisfy this equation and are related to the quantities L_a, $A_k/A_0\cdots$, $B_l/A_0\cdots$, by Eq. (15).

Then substituting for L_b, $\cdots$ above, one obtains

$$\left[L_a+\frac{\omega_{ab}}{\beta_{ab}}+\frac{l\omega}{\beta_{ab}}\right]\frac{B_{-l}}{A_0}=-\frac{A_{-l+1}}{A_0}-\frac{A_{-l-1}}{A_0},$$

$$\left[L_a+\frac{k\omega}{\beta_{ab}}\right]\frac{A_{-k}}{A_0}=-\frac{B_{-k+1}}{A_0}-\frac{B_{-k-1}}{A_0},$$

which is (8) again. Thus we have proved that the new set of quantities defined in Eq. (15) is a solution of (9) if L_a, $A_k/A_0\cdots$, $B_0/A_0\cdots$ satisfy (8).

APPENDIX 3.3a. SOLUTION OF EQ. (37) TO SECOND ORDER

Equation (37) can be written as

$$L_a=\frac{1}{L_a+\dfrac{\omega_{ab}-\omega}{\beta_{ab}}+\dfrac{\beta_{ab}}{2\omega}+\dfrac{\beta_{ab}^2}{4\omega_{ab}^2}L_a}+\frac{\beta_{ab}}{\omega_{ab}+\omega}.$$

Now clear of fractions and make use of the relations

$$\frac{1}{\omega}=\frac{1}{\omega_{ab}}+\frac{\omega_{ab}-\omega}{\omega_{ab}^2}+\cdots$$

$$\frac{1}{\omega_{ab}+\omega}=\frac{1}{2\omega_{ab}}+\frac{\omega_{ab}-\omega}{4\omega_{ab}^2}+\cdots$$

neglecting terms which are small when $(\omega_{ab}-\omega)/\omega_{ab}\ll 1$. We then have

$$L_a^2\left[1+\frac{\beta_{ab}}{2\omega_{ab}^2}\right]+\frac{\omega_{ab}-\omega}{\beta_{ab}}L_a=1+\frac{\omega_{ab}-\omega}{\omega_{ab}+\omega}+\frac{\beta_{ab}^2}{4\omega_{ab}^2}.$$

This quadratic equation has the solutions:

$$L_a=-\frac{\omega_{ab}-\omega}{2\beta_{ab}}+\frac{\beta_{ab}^2}{4\omega_{ab}^2}\frac{\omega_{ab}-\omega}{\beta_{ab}}\pm\frac{y}{2}\left[1-\frac{\beta_{ab}^2}{2y^2\omega_{ab}^2}\left(1+\frac{(\omega_{ab}-\omega)^2}{\beta_{ab}^2}\right)\right]^{\frac{1}{2}},$$

where

$$y=+\left[\frac{(\omega_{ab}-\omega)^2}{\beta_{ab}^2}+\frac{8\omega_{ab}}{\omega_{ab}+\omega}\right]^{\frac{1}{2}}.$$

After expanding the square root, this can be written:

$$L_a=\rho_\pm+\frac{\beta_{ab}^2}{4\omega_{ab}^2}\left[\frac{\omega_{ab}-\omega}{\beta_{ab}}\frac{1}{y}\mp\frac{(\omega_{ab}-\omega)^2}{\beta_{ab}^2 y}\right],$$

where $\rho_\pm$ are our previous solutions for L_a, as given by Eq. (33). The upper signs are to be used when $\omega<\omega_{ab}$ and the lower signs when $\omega\gtrsim\omega_{ab}$.

Substitute for y in this equation:

$$L_a=\rho_\pm\left[1-\frac{\dfrac{\beta_{ab}^2}{4\omega_{ab}^2}\left(1-2\dfrac{\omega_{ab}-\omega}{\beta_{ab}}\rho_\pm\right)}{2(\rho_\pm)^2+\dfrac{\omega_{ab}-\omega}{\beta_{ab}}\rho_\pm}\right].$$

Remembering that ρ satisfies Eq. (32) and substituting for $[(\omega_{ab}-\omega)/\beta_{ab}]\rho$, one obtains Eq. (38). In this step the terms $(\omega_{ab}-\omega)/(\omega_{ab}+\omega)(\beta_{ab}^2/4\omega_{ab}^2)\rho$ are neglected for they are always of higher than second order.

The correction term in (38) should be omitted when $2\rho^2-1$ becomes negative, for then its contribution becomes comparable to third-order terms which have already been omitted.

APPENDIX 3.4a. AMPLITUDE COEFFICIENTS NEAR RESONANCE

The other amplitude coefficients in the resonant region are:

$$\frac{B_{-1}}{A_0}=-\frac{A_{+1}}{B_0}=\frac{-1}{L_a+(\omega_{ab}+\omega)/\beta_{ab}+\cdots}=-\left[\frac{\beta_{ab}}{\omega_{ab}+\omega}-\frac{\beta_{ab}^2}{(\omega_{ab}+\omega)^2}L_a+\cdots\right].$$

Using the approximation for $\beta_{ab}/(\omega_{ab}+\omega)$ again leads to Eq. (43). Similarly:

$$\frac{A_{+2}}{A_0}=\frac{B_{-2}}{B_0}=\frac{-1}{L_a-2\omega/\beta_{ab}}\frac{B_{+1}}{A_0}=\frac{B_{+1}}{A_0}\left[\frac{\beta_{ab}}{2\omega}+\frac{\beta_{ab}^2}{4\omega^2}\rho\right],$$

$$\frac{A_{+2}}{A_0}=\frac{B_{+1}}{A_0}\left[\frac{\beta_{ab}}{2\omega_{ab}}+\frac{\beta_{ab}^2}{4\omega_{ab}^2}\left(\frac{2(\omega_{ab}-\omega)}{\beta_{ab}}+\rho\right)\right].$$

APPENDIX 4.2a. CALCULATION OF $(B_{+1})/A_0$ TO SECOND ORDER NEAR RESONANCE

We wish to employ Eq. (41) and obtain a more explicit solution for B_{+1}/A_0. First β_{ab}/K_{ab} must be calculated. Use Eq. (40) and expand the radical. Then, for $\omega \gtrless \omega_{ab}$

$$\frac{K_{ab}}{\beta_{ab}}=-1+\frac{\omega_{ab}-\omega}{2\beta_{ab}}-\frac{(\omega_{ab}-\omega)^2}{8\beta_{ab}^2}-\frac{\omega_{ab}-\omega}{2(\omega_{ab}+\omega)}.$$

By taking the reciprocal of this equation, one obtains:

$$\frac{\beta_{ab}}{K_{ab}}=-\left[1+\frac{\omega_{ab}-\omega}{2\beta_{ab}}+\frac{(\omega_{ab}-\omega)^2}{8\beta_{ab}^2}-\frac{\omega_{ab}+\omega}{4\omega_{ab}}\right].$$

Use this expression in Eq. (41) and also make use of the fact that $(2\rho^2-1)/(\rho^2+1)\approx\frac{1}{2}$ very near resonance. Then Eq. (50) follows.

Calculation of ω_0 to Second Order

Substitute Eqs. (50) and (51) into (49):

$$1+\frac{\beta_{ab}}{2\omega_{ab}}+\frac{\beta_{ab}^2}{8\omega_{ab}^2}+\frac{\beta_{cd}}{\omega_{cd}-\omega_{ab}}+\frac{\beta_{ab}\beta_{cd}}{2\omega_{ab}(\omega_{cd}-\omega_{ab})}+\frac{\beta_{ab}^2}{4\omega_{ab}^2}$$

$$+\frac{\beta_{ab}\beta_{cd}}{2\omega_{ab}(\omega_{cd}-\omega_{ab})}+\frac{\omega_{ab}-\omega_0}{\beta_{ab}}\left[\frac{1}{2}+\frac{\beta_{ab}}{4\omega_{ab}}\right.$$

$$\left.+\frac{\beta_{cd}}{2(\omega_{cd}-\omega_{ab})}\right]+\frac{(\omega_{ab}-\omega_0)^2}{8\beta_{ab}^2}$$

$$=1-\frac{\beta_{cd}}{\omega_{cd}-\omega_{ab}}-\frac{\beta_{ab}^2}{4\omega_{ab}^2}-\frac{\beta_{ab}\beta_{cd}}{2\omega_{ab}(\omega_{cd}-\omega_{ab})}+\frac{\beta_{ab}\beta_{cd}}{2\omega_{ab}(\omega_{cd}-\omega_{ab})}.$$

This simplifies to:

$$\frac{X^2}{8}+X\left[\frac{1}{2}+\frac{\beta_{ab}}{4\omega_{ab}}+\frac{\beta_{cd}}{2(\omega_{cd}-\omega_{ab})}\right]$$

$$=-\frac{\beta_{ab}}{2\omega_{ab}}-\frac{2\beta_{cd}}{\omega_{cd}-\omega_{ab}}-\frac{5}{8}\frac{\beta_{ab}^2}{\omega_{ab}^2}-\frac{3}{4}\frac{\beta_{ab}\beta_{cd}}{\omega_{ab}(\omega_{cd}+\omega_{ab})},$$

where

$$X=\frac{\omega_{ab}-\omega_0}{\beta_{ab}}.$$

By ignoring the X^2 term temporarily, one obtains:

$$X'=-\frac{\beta_{ab}}{\omega_{ab}}-\frac{4\beta_{cd}}{\omega_{cd}-\omega_{ab}}-\frac{3}{4}\frac{\beta_{ab}^2}{\omega_{ab}^2}+\frac{4\beta_{cd}^2}{(\omega_{cd}-\omega_{ab})^2}.$$

To second order:

$$(X')^2=\frac{\beta_{ab}^2}{\omega_{ab}^2}+\frac{16\beta_{cd}^2}{(\omega_{cd}-\omega_{ab})^2}+\frac{8\beta_{ab}\beta_{cd}}{\omega_{ab}(\omega_{cd}-\omega_{ab})}.$$

Now subtract $(X')^2/2$ from X',

$$X=-\frac{\beta_{ab}}{\omega_{ab}}-\frac{4\beta_{cd}}{\omega_{cd}-\omega_{ab}}-\frac{\beta_{ab}^2}{\omega_{ab}^2}-\frac{2\beta_{ab}\beta_{cd}}{\omega_{ab}(\omega_{cd}-\omega_{ab})},$$

from which Eq. (52) follows.

Electron Spin Coupling in Polyatomic Molecules*

R. S. HENDERSON†
Harvard University, Cambridge, Massachusetts
(Received July 6, 1955)

A formula to describe the splitting of rotational energy levels in polyatomic molecules, brought about by interaction of the magnetic field produced by rotation with electron spin and orbital angular momenta, is derived by perturbation methods on the basis of simplifying assumptions. The formula is then fitted to experimental data obtained from observations on fine structure in the spectrum of chlorine dioxide. The agreement is quite satisfactory.

INTRODUCTION

SOME observations upon molecular spectra[1] indicate that whenever there is a nonvanishing resultant of the electron spin moments within a polyatomic molecule, its energy levels exhibit a fine structure. In what follows we shall attribute this to the influence which the magnetic field induced by rotation must exert upon the electron spin moment, and upon the spin-orbit coupling, within the molecule.[2]

MOLECULAR MODEL

A molecule is pictured as a well-nigh rigid framework of atomic nuclei within which the lighter electrons carry out their rapid periodic evolutions; when the framework rotates, the electrons are carried along with it so that their motion relative to the nuclei is very little altered. Four kinds of angular momentum are then to be distinguished: (1) that of molecular rotation, (2) that of orbital motion of electrons, (3) that of spinning electrons, (4) that of spinning atomic nuclei.

We assume the following pattern of relationships among these momenta:

(a) Momenta of each type join to form resultants conveniently designated as **R** (molecular rotation), **L** (orbital motion of electrons), **S** (electron spin), and **I** (nuclear spin). These are taken to be independent of each other.

(b) Coupling among these resultants is of the sort commonly referred to as "Hund's Case b." Neither **R** nor **L** is a constant of the motion, but their resultant **N** is very nearly constant in magnitude and executes slow precession about an axis fixed in space. The electronic and nuclear spin moments are specified by the usual quantum numbers S and I. We shall restrict the present discussion to the case where $I=0$. The electron spin **S** is then assumed to join with **N** to form a final resultant **J**.

(c) Magnetic dipole interaction of these moments contributes a complex expression to the molecular Hamiltonian.[3,4] We proceed, however, on the simplifying assumptions that the magnetic field arising from rotation of the framework of nuclei can be taken as a linear combination of the vector components of **R** and that, despite the noncentral electric field to be expected in a molecule, the interaction energy can then be represented in reasonable approximation as $A'(\mathbf{L}\cdot\mathbf{S})+(a'\mathbf{R}_x+b'\mathbf{R}_y+c'\mathbf{R}_z)\cdot\mathbf{S}$.[5] Inasmuch as $\mathbf{R}=\mathbf{N}-\mathbf{L}$ and A' is much greater than a', b', or c',[6] this is very nearly the same as $A'(\mathbf{L}\cdot\mathbf{S})+(a'\mathbf{N}_x+b'\mathbf{N}_y+c'\mathbf{N}_z)\cdot\mathbf{S}$.

PERTURBATION PROBLEM

To get at the interplay between molecular rotation and electron spin, we now bracket the approximate spin coupling terms with that part of the Hamiltonian which describes rotation. This is done on the familiar hypothesis that rotation of the molecule is pretty much independent of other motions, but that the energy associated with spin coupling will depend upon it on account of magnetic dipole interactions involving **R**. Attention is accordingly to be fixed upon the expression

$$\frac{(N_x-L_x)^2}{2A}+\frac{(N_y-L_y)^2}{2B}+\frac{(N_z-L_z)^2}{2C}+A'(\mathbf{L}\cdot\mathbf{S})+(a'\mathbf{N}_x+b'\mathbf{N}_y+c'\mathbf{N}_z)\cdot\mathbf{S},$$

where x, y, z are the principal axes of the molecule and A, B, C the principal moments of inertia. This can be expanded as follows:

$$H_0=\frac{N_x^2}{2A}+\frac{N_y^2}{2B}+\frac{N_z^2}{2C};\quad H_0'=\frac{L_x^2}{2A}+\frac{L_y^2}{2B}+\frac{L_z^2}{2C};$$

$$H_1=-\left\{\frac{N_xL_x+L_xN_x}{2A}+\frac{N_yL_y+L_yN_y}{2B}+\frac{N_zL_z+L_zN_z}{2C}\right\};$$

$$H_2=A'(\mathbf{L}\cdot\mathbf{S});\quad H_3=(a'\mathbf{N}_x+b'\mathbf{N}_y+c'\mathbf{N}_z)\cdot\mathbf{S}.$$

* Part of a thesis submitted by R. S. Henderson to the Graduate School of Harvard University in partial fulfillment of the requirements for the Ph.D. degree. A summary was published as a Letter to the Editor: R. S. Henderson and J. H. Van Vleck, Phys. Rev. 74, 106 (1948).

† Now at Georgetown University, Washington, D. C.

[1] J. B. Coon, J. Chem. Phys. 14, 665 (1946).

[2] See G. C. Wick, Phys. Rev. 73, 51 (1948).

[3] J. H. Van Vleck, Revs. Modern Phys. 23, 216 (1951).

[4] G. R. Gunther-Mohr *et al.*, Phys. Rev. 94, 1191 (1954).

[5] See J. H. Van Vleck, *The Theory of Electric and Magnetic Susceptibilities* (Clarendon Press, Oxford, 1932), Sec. 40.

[6] The second-order terms are proportional to (electronic angular momentum)^2/(molecular moment of inertia)$(h\nu)\sim$(angular velocity of rotation)$h^2/h(h\nu)\sim$mass of electrons/mass of nuclei. The magnetic moment associated with **R** must bear a similar ratio to that associated with **L**, and hence first- and second-order terms in (4) must be of comparable magnitude. A rough

The term H_0' has no dependence either upon rotation or upon spin coupling and will henceforth be ignored.

The dominant term H_0 is recognized as the Hamiltonian of an asymmetric rotator in the absence of torques. Its matrix elements are most conveniently set down in a representation based on the wave functions of a symmetric rotator; these have been worked out by Wang[7] as follows:

$$\begin{aligned} H_0(NKM;NKM)&=\tfrac{1}{2}h^2c[GN(N+1)+K^2],\\ H_0(NK\pm 2M;NKM)&=\tfrac{1}{2}h^2cb[f(N,K\pm 1)], \end{aligned} \tag{1}$$

where $f(N,K)=-\frac{1}{2}[(N-K)(N-K+1)(N+K)(N+K+1)]^{\frac{1}{2}}$, $c=1/C-\frac{1}{2}(1/A+1/B)$, $G=(1/2c)(1/A+1/B)$, and b, well named the coefficient of asymmetry, is defined as $(1/2c)(1/A-1/B)$, axes having been chosen so that $-1\leqslant b\leqslant 0$. The symbol K stands for the quantum number typical of symmetric rotators and represents the quantized projection of $\mathbf{N}$ upon the axis of symmetry, while M represents the projection of $\mathbf{N}$ on the fixed axis about which it precesses.

The term H_1 containing cross products of $\mathbf{L}$ and $\mathbf{N}$ has its origin in the rotational distortion of electronic motion. To the extent that this may be neglected, the components of $\mathbf{L}$ along the body axes of the molecule appear in matrix form as $L_x(n';n)$, $L_y(n';n)$, $L_z(n';n)$, the letter n being used to denote the totality of electronic and vibrational quantum numbers. In a diatomic molecule, the two charged nuclei give rise to an electric field which is highly noncentral but axially symmetric, so that $\mathbf{L}$ may be presumed to have a quantized component along the axis of figure. But in a polyatomic molecule, no such field can be expected and all diagonal elements of $\mathbf{L}$ may be supposed to vanish.[8]

The vector $\mathbf{N}$ will have, in our representation based on the wave functions of the symmetric rotator, a component $N_z=Kh$. Furthermore the commutation relations among the components of $\mathbf{N}$ taken along the principal axes of the molecule are[9] $N_yN_z-N_zN_y=ihN_x$, etc. These relations differ only in the sign of i from those usually holding among angular momentum operators under these conditions.[3] One can make use of them in customary fashion[10] to work out the matrix elements of N_x and N_y. With suitable choice of phase, the result is

$$\begin{aligned}(N_y\pm iN_x)(NK\pm 1M;NKM)\\ =h[(N\mp K)(N\pm K+1)]^{\frac{1}{2}}.\end{aligned} \tag{2}$$

It is seen that within the limits of our assumptions the operators $\mathbf{N}$ and $\mathbf{L}$ are characterized by different quantum numbers. Their components therefore commute and the nonvanishing elements of H_1 emerge as

$$\begin{aligned} H_1(n'NKM;nNKM)&=-K(h/C)L_z(n';n),\\ H_1(n'NK\pm 1M;nNKM)&=\tfrac{1}{2}h(L_y/B\mp iL_x/A)(n';n)[(N\mp K)(N\pm K+1)]^{\frac{1}{2}}. \end{aligned} \tag{3}$$

That part of the molecular Hamiltonian which now concerns us is $H=H_0+(H_1+H_2+H_3)$, with the latter terms appearing in the role of a small perturbation upon H_0. Owing to the "off-diagonal" nature of $\mathbf{L}$, only H_3 among these perturbing terms contributes first-order corrections to H_0. Its elements diagonal in N may be set down as

$$(a'\mathbf{N}_x+b'\mathbf{N}_y+c'\mathbf{N}_z)\cdot\mathbf{N}[(\mathbf{N}\cdot\mathbf{S})/N(N+1)h^2]=(a'N_x^2+b'N_y^2+c'N_z^2)[(\mathbf{N}\cdot\mathbf{S})/N(N+1)h^2].$$

With the help of (2) above, we then find readily

$$H_3(NKM;NKM)=[N(N+1)(a'+b')+K^2(2c'-a'-b')]\frac{(\mathbf{N}\cdot\mathbf{S})}{2N(N+1)},$$

$$H_3(NK\pm 2M;NKM)=(a'-b')f(N,K\pm 1)\frac{(\mathbf{N}\cdot\mathbf{S})}{2N(N+1)}.$$

When this is transformed to a representation based on the wave functions of an asymmetric rotator, the resulting diagonal elements will constitute the first-order correction terms.

Second-order corrections of corresponding magnitude may also be turned up by applying to H a now well-known transformation[11] which has the effect of diagonalizing its matrix to a higher degree of approximation. The Hamil-

calculation shows that for light molecules this "dipole" correction may amount to as much as N cm^{-1}. See, e.g., N. F. Ramsey, Phys. Rev. **58**, 226 (1940).

[7] S. C. Wang, Phys. Rev. **34**, 243 (1929).

[8] Reference 5, Sec. 68.

[9] O. Klein, Z. Physik **58**, 730 (1929).

[10] See, e.g., E. V. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge University Press, Cambridge, 1935), p. 48.

[11] Termed a Van Vleck transformation. See E. C. Kemble, *The Fundamental Principles of Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1937), Sec. 48c.

tonian then takes the form $H=H_0+H_1+C_1+C_2+\cdots$ where

$$C_1(n''N''K''M'';nNKM)=\sum_{\substack{n'\neq n\\ N'K'M'}}\left[\frac{1}{E_0(n')-E_0(n)}-\frac{1}{E_0(n'')-E_0(n')}\right]$$
$$\times[H_1(n''N''K''M'';n'N'K'M')H_2(n'N'K'M';nNKM)$$
$$+H_2(n''N''K''M'';n'N'K'M')H_1(n'N'K'M';nNKM)],$$

$$C_2(n''N''K''M'';nNKM)=\sum_{\substack{n'\neq n\\ N'K'M'}}\left[\frac{1}{E_0(n')-E_0(n)}-\frac{1}{E_0(n'')-E_0(n')}\right]$$
$$\times[H_2(n''N''K''M'';n'N'K'M')H_2(n'N'K'M';nNKM)]$$

and E_0 stands for the unperturbed energies. We shall designate C_1 as the "dipole," C_2 the "pseudo-quadrupole" correction.

At this point it is well to recall that most polyatomic molecules have no axis of symmetry; eventually, therefore, we must carry out a further transformation from the present "symmetric top" representation to one based on the wave functions of an asymmetric rotator. Such a transformation has been investigated by Wang.[7] He finds the unitary transformation matrix $\|T\|$ to be in "step" form, lacking all elements which connect states of different n, N, or M. This means that in working out the diagonal elements of $\|T^{-1}\|\|C_1+C_2\|\|T\|$ which constitute our second-order correction terms, we shall find it unnecessary to consider elements of $\|C_1\|$ and $\|C_2\|$ connecting these states.

"DIPOLE" CORRECTION TERMS

The matrix elements of C_1 which we do need are to be obtained from those of H_1 and H_2. It is seen from Eqs. (3) that the matrix of H_1 is diagonal in N and M so that our task is reduced to that of finding those elements of H_2 in the symmetric top representation which connect states of the same N and M.

We begin by taking

$$H_2=A'(\mathbf{L}\cdot\mathbf{S})=A'(LS)P_1^0[\cos(L,S)]=A'[L_ZS_Z+\tfrac{1}{2}\{(L_X+iL_Y)(S_X-iS_Y)-(L_X-iL_Y)(S_X+iS_Y)\}],$$

where the three quantities L_Z, L_X+iL_Y, and L_X-iL_Y are referred to axes X, Y, and Z fixed in space. These quantities transform under a rotation of the reference axes in the same way as do N_Z, N_X+iN_Y, and N_X-iN_Y or any other triple formed in this way from the components of a vector. It may be shown from the theory of groups[12] that the matrix elements for one set of these quantities are proportional to the corresponding elements of another set, the factors of proportionality being independent of M. In the present case then

$$H_2(n'NK'MS;nNKMS)$$
$$=A'(\mathbf{L}\cdot\mathbf{S})(n'NK'MS;nNKMS)$$
$$=F(nn'NKK')A'(\mathbf{N}\cdot\mathbf{S})(NMS;NMS).$$

Matrix elements for the components of $\mathbf{L}$ in the fixed frame of reference may be expected to depend on the rotational quantum numbers even though, in the (xyz) frame mounted on the molecule, $\mathbf{L}$ has been assumed to be entirely decoupled from rotation. Reiche and Rademacher[13] have worked out these elements and confirmed this expectation.

The factors $F(nn'NKK')$ may now be determined explicitly by consideration of a special case; that in which $M=N$ is particularly suited to this purpose. Since L_X and L_Y have no matrix elements diagonal in M,[13] we obtain at once

$$A'(\mathbf{L}\cdot\mathbf{S})(n'NK'NS;nNKNS)$$
$$=A'L_Z(n'NK'N;nNKN)S_Z(S;S)$$
$$=F(nn'NKK')A'(\mathbf{N}\cdot\mathbf{S})(NNS;NNS)$$
$$=F(nn'NKK')A'N\hbar S_Z(S;S).$$

On substituting for L_Z from reference 13 we then find

$$F(nn'NKK)=\frac{KL_z(n';n)}{N(N+1)\hbar},\quad F(nn'NKK\pm1)=-[(N\mp K)(N\pm K+1)]^{\frac{1}{2}}\frac{(L_y\mp iL_z)(n';n)}{2N(N+1)\hbar}.$$

One may now proceed straightforwardly with the work of computing the matrix elements of C_1. Since the only "off-diagonal" elements of Wang's matrix $\|T\|$ are those connecting states specified by the quantum number K

[12] See C. Eckart, Revs. Modern Phys. 2, 305 (1930).
[13] F. Reiche and H. Rademacher, Z. Physik 41, 453 (1927).

with those specified by $K\pm2$, we need only the following:

$$C_1(nNKM;nNKM)=\frac{-2A'(\mathbf{N}\cdot\mathbf{S})(NMS;NMS)}{N(N+1)}\left\{K^2\left(\frac{2\gamma}{C}-\frac{\alpha}{A}-\frac{\beta}{B}\right)+N(N+1)\left(\frac{\alpha}{A}+\frac{\beta}{B}\right)-\frac{1}{2}iK\left(\frac{1}{B}-\frac{1}{A}\right)\delta_-\right\}$$

$$C_1(nNK\pm2M;nNKM)=\frac{2A'(\mathbf{N}\cdot\mathbf{S})(NMS;NMS)}{N(N+1)}\left\{f(N,K\pm1)\left[\left(\frac{\beta}{B}-\frac{\alpha}{A}\right)\mp\frac{1}{2}i\left(\frac{1}{B}+\frac{1}{A}\right)\delta_+\right]\right\},$$

where

$$\alpha=\sum_{n'\neq n}\frac{|L_x(n';n)|^2}{E_0(n')-E_0(n)};\quad \beta=\sum_{n'\neq n}\frac{|L_y(n';n)|^2}{E_0(n')-E_0(n)};\quad \gamma=\sum_{n'\neq n}\frac{|L_z(n';n)|^2}{E_0(n')-E_0(n)};$$

$$\delta_\pm=\sum_{n'\neq n}\frac{L_x(n';n)L_y^*(n';n)\pm L_x^*(n';n)L_y(n';n)}{E_0(n')-E_0(n)}$$

and the Hermitian properties of L_x, L_y, and L_z have been used throughout.

It is seen that δ_+ is the sum, δ_- the difference of complex conjugates. Furthermore, to the extent that coupling of $\mathbf{L}$ with other angular momenta may be disregarded, as we have done in supposing the matrix of $\mathbf{L}$ to lack diagonal elements, the matrix elements of L_x and L_y may be taken as pure imaginaries.[8] The products $L_x(n';n)L_y^*(n';n)$ are then real, so that δ_- vanishes while δ_+ is real.

Now the vector model adopted was one in which $\mathbf{N}$ and $\mathbf{S}$ were compounded to form a quantized resultant $\mathbf{J}$, whereas we have here, for reasons of convenience, been working with matrix elements which represent $\mathbf{N}$ as having a quantized component parallel to the fixed axis Z. A transformation from the $(nNKMS)$ representation to a $(nNKJS)$ one is therefore in order. It may be performed quite simply, for both C_1 and H_3 consist of factors dependent on quantum numbers common to both schemes of representation (which come through the transformation unaltered) multiplied by $(\mathbf{N}\cdot\mathbf{S})$ which in the new representation becomes $\frac{1}{2}\hbar^2[J(J+1)-N(N+1)-S(S+1)]$.

There remains the Wang transformation. Since C_1 is Hermitian, the purely imaginary terms containing δ_+ can contribute nothing to the diagonal elements of its transform. The rest of $\|C_1\|$ together with $\|H_3\|$ may be arranged as the sum of a matrix proportional to that of H_0 [see Eq. (1)] and one which is diagonal in the symmetrical top system of representation. The first is rendered diagonal by Wang's transformation and its contribution to our formula for fine structure is therefore the expression for its characteristic values which work out as

$$\left[4A'\left(\frac{\beta}{B}-\frac{\alpha}{A}\right)-(b'-a')\right]\left[\frac{J(J+1)-N(N+1)-S(S+1)}{2N(N+1)bc}\right]W(K)$$

with $W(K)$ standing for the energy levels of an asymmetric rotator characterized by the parameters b, c, and G, Principal elements for the transform of the second are readily computed.

Upon assembling our results, both first and second order, we have the "dipole" correction in its final form. namely[6]

$$\frac{\hbar^2[J(J+1)-N(N+1)-S(S+1)]}{4N(N+1)}\left\{\left[8A'\left(\frac{\beta}{B}-\frac{\alpha}{A}\right)-2(b'-a')\right]\frac{W(K)}{cb\hbar^2}\right.$$

$$+\left[(a'+b')+(G/b)(b'-a')-4A'\left(\frac{\alpha}{A}+\frac{\beta}{B}\right)-4A'(G/b)\left(\frac{\beta}{B}-\frac{\alpha}{A}\right)\right]N(N+1)$$

$$\left.+\left[2c'-a'-b'+(1/b)(b'-a')-4A'\left(\frac{2\gamma}{C}-\frac{\alpha}{A}-\frac{\beta}{B}\right)-4A'(1/b)\left(\frac{\beta}{B}-\frac{\alpha}{A}\right)\right]\sum_{K'}T^2(K';K)K'^2\right\}. \quad (4)$$

"PSEUDO-QUADRUPOLE" CORRECTION TERMS

Once again the simple form of $\|T\|$ makes it unnecessary to consider elements of $\|C_2\|$ other than those which connect states of the same n, N, and M. The following alone are needed:

$$C_2(nNK''M;nNKM)=\sum_{n'\neq n}\frac{2(A')^2}{E_0(n')-E_0(n)}\left\{\sum_{N'K'M'}(\mathbf{L}\cdot\mathbf{S})(nNK''M;n'N'K'M')(\mathbf{L}\cdot\mathbf{S})(n'N'K'M';nNKM)\right\},$$

with $K''=K$, $K\pm2$.

The sum over the rotational quantum numbers is equivalent to

$$[L^2S^2\cos^2(L,S)](n'NK''M;nNKM)=[(L^2S^2)\{\tfrac{2}{3}P_2^0[\cos(L,S)]+\tfrac{1}{3}\}](n'NK''M;nNKM).$$

The term $\frac{1}{3}(L^2S^2)$ does not depend on the rotational quantum numbers and need receive no further notice. With the help of the addition theorem for Legendre Polynomials, the remaining term may be recast as

$$\left[\tfrac{2}{3}\sum_{s=-2}^{s=2}\Pi_2^s(L)\pi_2^{-s}(S)\right](nNK''M;nNKM)^{14}$$

where $\pi_2^{\pm2}(Q)=\frac{3}{8}^{\frac{1}{2}}(Q_X\pm iQ_Y)^2$, $\pi_2^{\pm1}(Q)=\frac{3}{8}^{\frac{1}{2}}[Q_Z(Q_X\pm iQ_Y)+(Q_X\pm iQ_Y)Q_Z]$, $\pi_2^0(Q)=[\frac{3}{2}Q_Z^2-\frac{1}{2}Q_X^2-\frac{1}{2}Q_Y^2-\frac{1}{2}Q_Z^2]$.

These five quantities form a manifold invariant under rotation of the fixed (XYZ) axes.[15] Hence, as in the preceding section, we may borrow from the theory of groups to write $[\pi_2^s(L)](NK''M;NKM)=F(NK''K)\times[\pi_2^s(N)](NM;NM)$, where F is independent of M and s[16] and may be determined from a consideration of the special case in which $M=N$. Specifically, $[\pi_2^0(L)](NK''N;NKN)=F(NK''K)[\pi_2^0(N)](NN;NN)$, whence

$$F(NK''K)=\frac{3\{\sum_{n'N'K'}L_Z(nNK''N;n'N'K'N)L_Z(n'N'K'N;nNKN)\}-|L(n';n)|^2}{\hbar^2[3N^2-N(N+1)]}.$$

Matrix elements for the components of $\mathbf{L}$ in this expression may be entered from the list in reference 13.

We now have

$$C_2(nNK''M;nNKM)=\sum_{n'\neq n}\frac{(4/3)A'^2}{E_0(n')-E_0(n)}F(NK''K)\left[\sum_{s=-2}^{s=2}\pi_2^s(N)\pi_2^{-s}(S)\right](NM;NM).$$

The next step is to change over from the $(nNKMS)$ representation to a $(nNKJS)$ one. Such a change will not affect the factor $F(NK''K)$. As for the sum over s, Kramers has shown that it is transformed to diagonal form on passing to the new representation and he has calculated its characteristic values as $[\frac{3}{8}C(C+1)-\frac{1}{2}N(N+1)\times S(S+1)]\hbar^4$ where $C=[J(J+1)-N(N+1)-S(S+1)]$.[16]

Finally, Wang's transformation may be applied as before to obtain the desired correction formula

$$\frac{2(A'\hbar)^2[\frac{3}{8}C(C+1)-\frac{1}{2}N(N+1)S(S+1)]}{N(N+1)(2N-1)(2N+3)}\left\{\frac{-2(\beta-\alpha)}{bc\hbar^2}W(K)+[\tfrac{3}{2}(\alpha+\beta-2\gamma)+(G/b)(\beta-\alpha)]N(N+1)\right.$$
$$\left.-[2(\alpha+\beta-2\gamma)-(1/b)(\beta-\alpha)]\sum_{K'}T^2(K';K)K'^2\right\}. \quad (5)$$

The justification is now apparent for the names given to these two terms which together describe fine structure in the rotational energy levels of a polyatomic molecule. One has its origin in terms connecting the spin coupling energy with rotational distortion of electronic motion. It contains the quantum mechanical average of $\cos(N,S)$ and so resembles the Landé expression typical of dipole coupling. The other, proportional to the square of the spin-orbit coupling energy, stems from second order effects of magnetic dipole coupling, but it contains the average of $\cos^2(N,S)$ and bears, in consequence, a strong resemblance to expressions commonly obtained in calculations on quadrupole coupling.

ADAPTATION OF THE FORMULAS TO THE CASE OF NEARLY COMPLETE SYMMETRY

Both correction terms depend on the quantities $W(K)$ and $T(K';K)$ for which general formulas are lacking. They may be worked out more explicitly, however, in the important special case where two of the principal moments of inertia of the molecule are nearly equal in magnitude while differing considerably from the third. In this event, b is a small fraction whose square may be neglected, while the energy levels $W(K)$ will differ only slightly from those of the symmetric rotator. An exceptional case occurs when $K=\pm1$ and this must be treated separately; otherwise one may readily work out Wang's secular problem[7] to this degree of approximation with the following results:[17]

$$W(K)=\tfrac{1}{2}ch^2[GN(N+1)+K^2-\tfrac{1}{4}b^2U],$$
$$\sum_{K'}T^2(K';K)K'^2=K^2+\tfrac{1}{4}b^2U,$$

where

$$U=\left\{\frac{[f(N,K+1)]^2}{1+K}+\frac{[f(N,K-1)]^2}{1-K}\right\}.$$

Substitution of these expressions into formulas (4) and (5) gives us the explicit correction terms for which we

[14] H. B. G. Casimir, *On the Interaction between Atomic Nuclei and Electrons* (De Erven f. Bohn, Haarlem, 1936), pp. 12 and 13.

[15] E. Wigner, *Gruppentheorie* (Friedrich Vieweg & Sohn, Braunschweig, 1931), Chap. XV.

[16] H. A. Kramers, Proc. Roy. Acad. Amsterdam **34**, 965 (1931).

[17] E. E. Witmer, Proc. Natl. Acad. Sci. U. S. **13**, 60 (1927).

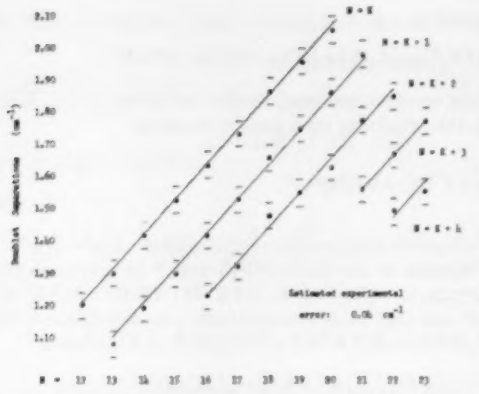


FIG. 1. Spin doublets in absorption spectra of chlorine dioxide.

seek. They turn out to be precisely similar to those obtained by Knight and Feld[18] in treating a closely related problem.

Higher order approximations for the effects of asymmetry may be calculated by extension of the foregoing methods.

FINE STRUCTURE IN THE ROTATIONAL ENERGY LEVELS OF CHLORINE DIOXIDE

Coon[1] has carried out a series of careful measurements on five absorption bands of ClO_2. These lie in the ultraviolet region between 4000 and 4700 angstroms, and each contains a wealth of rotational structure which fortunately is not obscured by overlapping. Coon found it possible to sort out some hundreds of individual lines and to classify them according to the "rotational transitions" from which they stemmed. Members of the so-called qQ branches, arising from transitions between states having the same rotational quantum numbers, revealed themselves clearly as doublets. This was attributed to spin coupling since the ClO_2 molecule contains an odd number (33) of electrons and has probably a total spin quantum number S equal to $\frac{1}{2}$.

Now this molecule is known to be of the "bent" triatomic type. Coon calculated its principal moments of inertia as $A=110\times10^{-40}$ g cm^2, $B=95\times10^{-40}$ g cm^2, $C=15\times10^{-40}$ g cm^2. The coefficient of asymmetry b is then only about 0.013. Indeed, the degree of asymmetry is so slight that the quantum number K retains the status of a "pretty good quantum number" serving to classify the $(2N+1)$ rotational states consistent with a given N and J.

Here is a case to which the present theory should apply. With $S=\frac{1}{2}$, the "pseudo-quadrupole" correction term vanishes. But since the moments of inertia are small while the values of N and K turn out to be large, one can expect the "dipole" expression of (4) to represent an effect of observable magnitude. This is approximately

$$\left[\lambda\frac{K^2}{N(N+1)}-\mu-\frac{\eta U}{N(N+1)}\right][J(J+1)-N(N+1)-S(S+1)],$$

where

$$\lambda=\left[\tfrac{1}{4}(2c'-a'-b')-\left(\frac{2\gamma}{C}-\frac{\alpha}{A}-\frac{\beta}{B}\right)A'\right]\hbar^2;\quad \mu=\left[\left(\frac{\alpha}{A}+\frac{\beta}{B}\right)A'-\tfrac{1}{4}(a'+b')\right]\hbar^2;\quad \eta=\tfrac{1}{2}b\left[\left(\frac{\beta}{B}-\frac{\alpha}{A}\right)A'-\tfrac{1}{4}(b'-a')\right]\hbar^2.$$

The quantum number J can have two values, $N+\frac{1}{2}$ and $N-\frac{1}{2}$. Let us call the corresponding energy levels D' and D''. It is not certain which of these is the greater since the quantities λ, μ, and η defy computation. The usual selection rules permit transitions in which $\Delta J=0,\pm1$; only those corresponding to $\Delta J=0$ are observed, intensities for the others being much weaker. The interval separating the qQ doublets then works out as follows:

$$\Delta\tilde{\nu}=(1/hc)[(D_{n_1}'-D_{n_2}')-(D_{n_1}''-D_{n_2}'')]=(1/hc)[(D_{n_1}'-D_{n_1}'')-(D_{n_2}'-D_{n_2}'')]$$
$$=\frac{(2N+1)}{hc}\left\{\frac{\Delta\lambda K^2}{N(N+1)}-\Delta\mu-\frac{\Delta\eta U}{N(N+1)}\right\},$$

where the subscripts n refer to the electronic states. Since the quantities a', b', and c' do not depend on n, they must cancel out in this final result.

The formula just obtained compares well with the empirical one

$$\Delta\tilde{\nu}=q(2K-N)-r(v-1)$$

given by Coon. For values of N and K large in comparison with their difference (the situation with which Coon was working) the leading term above becomes approximately

$$\Delta\lambda(2K^2/hcN)=(2\Delta\lambda/hcN)[N-(N-K)]^2$$
$$\doteq(2\Delta\lambda/hcN)[N^2-2N(N-K)]=(2\Delta\lambda/hc)(2K-N)$$

as in the empirical formula. The second part of the theoretical expression differs from that of Coon in showing a dependence on N, but it turns out to be small.

The agreement of the theoretical expression with the experimental data is good. (See Fig. 1.) A fit was

[18] G. B. Knight and B. T. Feld, Phys. Rev. **74**, 354(A) (1948).

obtained by the method of least squares using the doublet intervals in the third band where Coon considered his measurements to be most accurate. This gave $(\Delta\lambda/hc)=0.057$ cm^{-1}; $(\Delta\mu/hc)=0.0038$ cm^{-1}; $(\Delta\eta/hc)=0.0001$ to 0.0002 cm^{-1}. The constant $\Delta\eta$ is minute as one might anticipate from the fact that it is proportional to b and to $(\beta/B-\alpha/A)$, both of which vanish when the molecule has an axis of symmetry. Indeed, the results would not be altered appreciably if the term containing $\Delta\eta$ were omitted, that is, if the asymmetry of the molecule were disregarded entirely. It should be stated, however, that this term, tiny as it is, yet acts uniformly in such a direction as to improve the agreement between theory and experiment.

FINE STRUCTURE RESULTING FROM COUPLING OF NUCLEAR SPINS

There seems to be no reason why precisely similar results should not also apply when $S=0$ and nuclear spin holds the center of interest. An attempt to use them in describing fine structure observed in the inversion spectrum of $N^{14}H_3$ [19] has now been superseded by the excellent work of Gunther-Mohr.[4]

ACKNOWLEDGMENTS

My best thanks are due Professor J. H. Van Vleck under whose direction this work was done and whose ideas form the basis of it.

[19] R. S. Henderson, Phys. Rev. **74**, 107 (1948).

PHYSICAL REVIEW VOLUME 100, NUMBER 2 OCTOBER 15, 1955

Electron Loss Process in the Hydrogen Afterglow*

KARL-B. PERSSON† AND SANBORN C. BROWN
Department of Physics and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received May 19, 1955)

Electron loss processes in the afterglow of very pure hydrogen have been studied by the microwave technique. The conclusion is drawn that within the error of the experiment, no electron-ion recombination is observed, and that high-mode diffusion is very evident. By properly accounting for the high diffusion modes, the results are given as $D_a p=700\pm50$ cm^2-mm Hg sec^{-1} with the electron-ion recombination less than 3×10^{-6} cm^{-3} sec^{-1}.

ELECTRON loss processes in the hydrogen afterglow have been studied by the microwave method by a number of workers. Although the results are related, no agreement has been achieved up to the present time as to the actual loss processes in hydrogen.[1–3] In view of the result of the work presented here, it is possible to understand the disagreement between the previous investigations of the hydrogen loss processes. The main result is that, if sufficiently pure hydrogen is used, the loss process for the electrons in the ranges covered by previous investigators, as well as by us, is explained by an ambipolar diffusion mechanism. If the electron-ion recombination is present, this loss process is so small that it cannot be measured with methods available at present.

MICROWAVE METHOD

The microwave method measures the frequency shift, Δf, of a resonant cavity containing the electron-ion plasma, relative to the resonant frequency, f, of the same cavity without the plasma. The relation between the frequency shift and the average electron density, $\langle n\rangle$, is

$$\frac{\Delta f}{f}=\frac{1}{2}\frac{1}{1+(\nu_m/\omega)^2}\frac{\langle n\rangle}{n_p}, \tag{1}$$

where the average electron density $\langle n\rangle$ is given by

$$\langle n\rangle=\int_V nE^2dv\Big/\int_V E^2dv. \tag{2}$$

The quantity $n_p=\epsilon_0 m\omega^2/e^2$, E is the electric field of the probing microwave signal, ν_m is the momentum transfer collision frequency per electron (ν_m is proportional to the pressure), ω is the applied radian frequency of the probing signal, and V is the volume of the microwave cavity. The linear relationship between the frequency shift and the average electron density holds only as long as the polarization field in the plasma is small compared with the applied microwave field. The error in linearity is less than 1 percent if $n\leq3\times10^8[1+(\nu_m/\omega)^2]$ cm^{-3}.

To interpret the measurements of the microwave method, it is necessary to recognize the fact that only an average electron density is measured. Neglecting the influence of the shape of the microwave field, the microwave method measures the total number of the

* This work was supported in part by the Signal Corps; the Office of Scientific Research, Air Research and Development Command; and the Office of Naval Research.

† Now at Research Laboratory, General Electric Company, Schenectady, New York.

[1] M. A. Biondi and S. C. Brown, Phys. Rev. **76**, 1697 (1949).

[2] J. M. Richardson and R. B. Holt, Phys. Rev. **81**, 153 (1951).

[3] L. J. Varnerin, Jr., Phys. Rev. **84**, 563 (1951).

free electrons present in the cavity. Therefore, whatever the loss process is, the rate of loss of electrons refers to the total number of the electrons present. For a proper interpretation of the measurements it is necessary to know the electron density distribution. The three-dimensional electron density distribution, as well as the probing microwave mode, can be split into one symmetric and one antisymmetric part for each dimension. The microwave method using a symmetric mode measures only the symmetric part of the electron density distribution.

AMBIPOLAR DIFFUSION

We restrict our discussion to those times when the electrons can be considered in thermal equilibrium with the gas and when the space-charge-influenced diffusion is the dominant loss mechanism. The decay of the electron and ion densities is governed by the following equations:

$$\Gamma_- = -D_-\nabla n_- - \mu_- \mathbf{E} n_-,\ \Gamma_+ = -D_+\nabla n_+ + \mu_+ \mathbf{E} n_+,$$
$$\partial n_-/\partial t + \nabla\cdot\Gamma_- = 0;\ \partial n/\partial t + \nabla\cdot\Gamma_+ = 0,$$
$$\nabla\cdot\mathbf{E} = e(n_+ - n_-)/\epsilon_0. \quad (3)$$

Here Γ is the particle current density, D is the diffusion coefficient, μ the mobility; the subscripts indicate positive ions and electrons. A convenient system of proper variables for this equation system is: $t/p\Lambda^2$, x/Λ, n/Λ^2, $p\Lambda^2\Gamma$, and $E\Lambda$, where Λ is the fundamental diffusion length. The solution to this equation system can be written in the general form $n\Lambda^2 = f[(\mathbf{r}/\Lambda), (t/p\Lambda^2)]$, where $\mathbf{r} = f(x,y,z)$. Unfortunately this function is not separable in terms of the variables $\mathbf{r}/\Lambda$ and $t/p\Lambda^2$. Since the microwave method measures an average electron density the averaging process has to be applied to these equations. The general result can be written in the form $\Lambda^2\langle n\rangle = g(t/p\Lambda^2)$. This solution is rigorous when the ion and the electron temperatures are independent of the coordinates. This form of the solution is also independent of the magnitude of the space charge. The function, g, depends upon the shape of the probing microwave field and the initial spatial distribution.

In the ambipolar limit, the equation system (2) may be reduced to the following equation which governs the decay of both the electron and the ion densities

$$\partial n/\partial t = D_a\nabla^2 n. \quad (4)$$

The criterion for ambipolar diffusion described by Eq. (4) is $n > n_+ - n_- \approx \epsilon_0(kT/e^2\Lambda^2)$. The diffusion length used in this experiment was of the order of one centimeter, and at room temperature this criterion can be written as $n > 10^8\ \mathrm{cm}^{-3}$.

In the experiment to be described, the gas is contained in a cubical quartz bottle placed at the center of the microwave resonator. The general solution for the average electron density in the ambipolar limit is simple to derive but lengthy to write down. For the one-dimensional case the general solution of the electron spatial density distribution can be written as

$$n\sum_i A_i \exp[-(i\pi/L)^2 D_a t]\cos(i\pi x/L),$$
$$i = 1,3,5,7,\cdots. \quad (5)$$

The first term of this series is the fundamental diffusion mode and the higher terms are therefore properly called higher diffusion modes. In the one-dimensional case, the ratios between the time constant for the decay of the fundamental mode and the higher modes are so large (the minimum ratio is 9) that it is customary to neglect the higher modes. In the three-dimensional case, the general solution consists of the product of three series like (4). The time constant for the fundamental diffusion mode is $T_1 = L^2/3\pi^2 D_a$, where L is the side of the quartz bottle. The ratios between this time constant and those of the higher diffusion modes are

$$T_1/T_2 = 3.67;\ T_1/T_3 = 6.33;\ T_1/T_4 = 9.$$

Since these ratios are relatively small, and the amount of higher diffusion modes present may be large, these higher diffusion modes may be present to a measurable extent rather late in the postdischarge period. The initial electron spatial distribution may vary widely and it becomes impossible to recognize the ambipolar diffusion process from the time dependence of the decay curve alone. The characteristic feature of the ordinary diffusion process, as well as of the space-charge-influenced diffusion, is that by using the proper time variable $t/p\Lambda^2$, the time dependence of the average electron density becomes independent of the pressure p, the fundamental diffusion length Λ, and the absolute value of the average electron density $\langle n\rangle$, for any given initial spatial electron distribution.

ELECTRON ATTACHMENT SUPERIMPOSED ON AMBIPOLAR DIFFUSION

The difficult feature of experiments of this kind is the problem of purity of the gas. The most serious impurities are those that form negative ions; they are oxygen and water vapor. The probability that an electron will attach to one of these molecules at room temperature is around 10^{-3}. In the ambipolar diffusion range and when attachment is present, the decay of the electron density is governed by the differential equation

$$\partial n/\partial t = D_a\nabla^2 n - hkNn. \quad (6)$$

In the pure gas, the number of hydrogen atoms present is very small and we neglect the formation of negative hydrogen ions. The number density, N, is therefore the number density of the impurities; h is the corresponding probability for attachment; kN is the rate of collisions per electron with the impurities.

Assuming that the impurity desnity is so large that variations in N caused by the formation of negative ions can be neglected, the solution for average electron

density takes the form

$$\langle n\rangle = K e^{-hkNt} g(D_a t/\Lambda^2), \tag{7}$$

where $g(D_a t/\Lambda^2)$ is the solution with no impurities present. Later in the afterglow when the higher diffusion modes have decreased to a negligible value relative to the fundamental diffusion mode, the average electron density becomes

$$\langle n\rangle = K \exp\{-[hkN + (D_a p/p\Lambda^2)]t\}. \tag{8}$$

Attachment losses are thus dominant in the late afterglow when $hkN \gg D_a p/p\Lambda$. The $(D_a p)$ for hydrogen, as measured in this experiment, is 700 cm²—mm Hg/sec; $\Lambda^2 = 1$ cm². The collision frequency between electrons and the impurities can be estimated to be of the order of $10^{10}\ p_i$/sec, where p_i is the partial pressure of the impurities. The criterion for dominant attachment losses can then be written $p_i/p \gg 10^{-4}/p^2$, where p is the pressure of the hydrogen. The criterion for ambipolar diffusion to be dominant (to an error of less than one percent) is $p_i/p \ll 10^{-6}/p^2$. This indicates that extreme purity is required in order to measure ambipolar diffusion late in the afterglow. At a pressure of 10 mm Hg, an impurity of less than one part in 10^8 of hydrogen is necessary if the attachment process is to be neglected.

If the surfaces of the container have been cleaned during the discharge period, so that virtually no impurities are left on them, these surfaces will easily reabsorb impurities when the discharge is over. It is therefore possible that impurities present in the afterglow are lost to the walls by diffusion. The influence of the impurities can be seen by writing

$$\begin{aligned} \partial n/\partial t &= -[D_a/\Lambda^2 + hkN]n, \\ \partial N/\partial t &= -[D_i/\Lambda^2 + hkn]N, \end{aligned} \tag{9}$$

where D_i is the diffusion coefficient of the impurities. When we put numbers into these equations we find that the diffusion loss of the impurities is much larger than the rate at which the impurities are transformed into negative ions. Thus the term containing the electron density in the second equation of (9) may be neglected, and the rate of loss of the electron is determined by

$$\frac{\partial \ln n}{\partial (t/p\Lambda^2)} = -\left[D_a p + p\Lambda^2 N_0 hk \exp\left(-\frac{D_i p t}{p\Lambda^2}\right)\right]. \tag{10}$$

This equation shows that the attachment loss may be dominant in the early afterglow and negligible in the late postdischarge period. If the impurities originate on the walls of the container they must also leave the walls during the discharge pulse in a diffusion process. The product $p\Lambda^2 N_0$ is then essentially independent of p and Λ, provided that all other experimental parameters are kept constant. The process proposed here will then behave as a diffusion process as far as dependence on t, p, and Λ goes, although the attachment may be the dominant loss process of the electrons in the early afterglow. If a monomolecular layer of impurities is released from the walls during the discharge period in a container of the size used in this experiment (a cube with the side 5.5 cm), the partial pressure of the impurities would be of the order of 10^{-1} to 10^{-3} mm Hg. This quantity of impurities is sufficient to give dominant attachment loss process when the hydrogen pressure is larger than 10^{-1} mm Hg. It is thus not sufficient to have pure gas; the container must be cleaned with extreme care.

THE APPARATUS

A resonant cavity was built in the form of a parallelepiped. The three slightly different lengths of sides allowed three different resonant wavelengths for the three fundamental modes of the cavity. These wavelengths were 9.5 cm, 10.0 cm, and 10.5 cm. One mode was used by a power signal creating the plasma while the others could be used for microwave probing signals. A weak frequency-modulated probing signal was transmitted through the resonant cavity in parallel with a cavity wavemeter. After passing through the two cavities, the probing signal was rectified and observed as a function of time on an oscilloscope screen. The input and output couplings to the probing modes were so small that they had no measurable effect on the resonant frequency of that mode.

A photomultiplier and a slit system was moved along one of the walls of the cavity. The signal from the photomultiplier, displayed on an oscilloscope screen, reproduced the distribution of light from the plasma perpendicular to the wall at any chosen time. The light distribution at the end of the discharge pulse was used as a qualitative measure of the initial spatial electron distribution of the decay period.

The discharge plasma was contained in a cubical quartz bottle placed in the center of the microwave cavity. The bottle was connected to a vacuum system and a hydrogen supply. The influence of the fundamental diffusion length was found by using two quartz bottles of different sizes. The discharge pulses were repeated at a rate of 10–60 per second, and the pulse length could be varied from 50 msec to 0.5 μsec.

The vacuum system could reach a holding vacuum of 10^{-10} mm Hg after 24 hours. The hydrogen supply consisted of a quartz tube containing uranium hydride and a high-vacuum metal valve. The uranium hydride was made in the following way. The quartz tube was filled with uranium metal filings and baked at 800°C until a holding vacuum of less than 10^{-9} mm Hg was reached. Hydrogen was then let into the system and uranium hydride formed at about 100°C. When the hydride was heated to about 250°C, the hydrogen was released.

Contamination of the hydrogen during the discharge was a most serious problem even with the best baking,

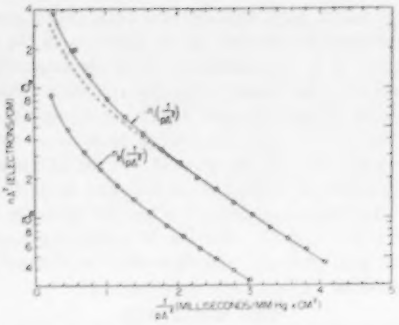


FIG. 1. The decay of electron density in hydrogen.

discharge cleaning, and vacuum techniques. The hydrogen could be shown to react with the quartz bottle during the discharge pulse. To avoid this reaction, the shortest possible discharge pulse, 0.5 μsec, was used.

MEASUREMENTS

Since it was found during the course of the investigation that impurities were produced during the discharge period, we were forced to use a short pulse length to achieve satisfactory purity. These short pulse lengths produced asymmetric spatial distributions of the discharge. In order to cover as large a pressure range as possible without changing the spatial distribution of the discharge, we found it experimentally necessary to use a very asymmetric discharge where high diffusion modes were prominent. However, if the diffusion process is the only loss mechanism, curves plotted against the proper variables for diffusion, $\langle n \rangle \Lambda^2$ and $t/p\Lambda^2$ should have the same shape, independent of the initial electron density, as long as the initial spatial distribution is held constnat. This is illustrated in Fig. 1. When these two curves are superimposed by sliding one of them in the vertical direction, they coincide within experimental error.

The initial spatial distribution of the electron density in the postdischarge period could be changed by varying the coupling of the input power. Figure 2 shows that when the initial distribution is very asymmetric the density decays very rapidly in the early afterglow, compared with the case when the initial distribution is more uniform. In the late afterglow, the two decay curves are parallel. This behavior is expected if higher diffusion modes are present. A higher density at the walls of the container would put a greater electron density in the higher diffusion modes than would be the case for a uniform distribution. Since the higher modes decay faster than the fundamental, this would explain the behavior of the curves. When the electron density has decayed to a point where only the fundamental diffusion mode is present, the density decay should be independent of the initial conditions, as illustrated.

Typical of the range of data covered by this experiment are the curves of Fig. 3. The initial spatial distribution is the same for all curves, although the initial densities vary from pressure to pressure. Since the data are plotted in terms of the proper variables for diffusion, $\langle n \rangle \Lambda^2$ and $t/p\Lambda^2$, if the data fit the ambipolar diffusion theory they should coincide when displaced vertically. The result of such normalization is shown in Fig. 4. Data were taken for other initial spatial distributions; for each constant distribution they exhibited a similar good fit with the diffusion variables.

CONCLUSION

The measurements shown above demonstrate that the diffusion mechanism is the dominant loss process of the electrons in the afterglow of pure hydrogen. Our discussion of the attachment loss process and impurities originating on the walls of the quartz bottle, as well as the necessity of using short discharge pulses to obtain purity, indicate that an attachment process may have been present in the early afterglow. The presence of higher diffusion modes made it impossible to establish this process in a more positive manner. Measurements of small amounts of impurities were quite outside the scope of this investigation. Fortunately, the result of the measurements in the late afterglow enabled us

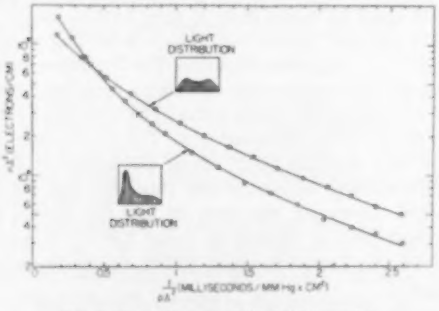


FIG. 2. The influence of the initial spatial distribution on the decay of electron density.

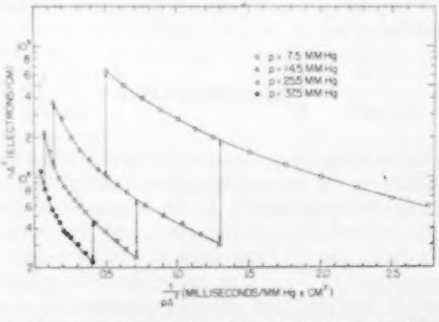


FIG. 3. Experimental runs of electron density decay at different pressures with constant initial electron distribution.

to draw definite conclusions. It can be seen from Fig. 4 that it takes considerable time for the higher diffusion modes, and possibly the attachment process, to disappear and leave the fundamental diffusion mode alone. In some cases, this does not occur until the average electron density is too small to be measured. It is therefore convenient to define an effective $D_a p$ as

$$(D_a p)_e = \frac{\partial \ln(\langle n \rangle \Lambda^2)}{\partial (t/p\Lambda^2)}.$$

This can be determined from the slope of the decay curves. It is plotted in Fig. 5. It is obvious from this figure that the effective diffusion coefficient multiplied by the pressure is not constant in the presence of the higher diffusion modes and possibly the attachment process. The value of $(D_a p)_e$ when only the fundamental diffusion mode is present can be taken as the true $D_a p$. For pure hydrogen this experiment gives $D_a p = 700 \pm 50$ cm²-mm Hg sec⁻¹.

Although it has been stated that decay curves taken at different initial electron densities and with the same spatial distribution have the same slope within experimental error, they do not coincide exactly. The dotted line on Fig. 1 illustrates the degree of departure allowable experimentally when the lower curve is displaced to coincide with the upper one. This deviation from similarity in shape could be attributed to other loss mechanisms than diffusion. If significance is attributed to this deviation (which is within experimental error) and it is attributed to recombination, we conclude that this recombination is less than 3×10^{-8} cm^{-3} sec^{-1}.

COMPARISON WITH PREVIOUS RESULTS

In 1949, Biondi and Brown[1] surveyed the electron-ion recombination losses in a number of gases and reported their findings in hydrogen to yield a recombination coefficient of 2.5×10^{-6} cm³/sec, independent of the pressure between 3 and 12 mm Hg. They used "spectroscopically pure" gas and relatively long discharge pulses of 250 μsec. Richardson and Holt[2] subsequently studied the same problem and found a recombination coefficient of 6.2×10^{-6} cm³/sec, independent of the pressure between 7 and 20 mm Hg, but in the lower pressure region rising from 2×10^{6} at 1 mm Hg to 6.2×10^{6} at 7 mm Hg. They also used "spectroscopically pure" gas but a smaller discharge pulse of 10 μsec. Varnerin[3] reported his results for hydrogen recombination to be somewhat similar. He found a constant recombination coefficient at pressures between 30 and 50 mm Hg, but a changing recombination coefficient below this pressure rising from 3×10^{-7} at 3 mm Hg to 2.5×10^{-6} at 30 mm Hg. Varnerin paid much more attention to purity, diffusing his hydrogen through hot palladium and baking his discharge tube thoroughly to outgas it. He used a discharge pulse of 1 μsec.

All of these determinations fall into the following pattern: The more carefully the purity of the gas is controlled, the lower the coefficient. At higher pressures (where collisions with impurities are more likely) the coefficient is higher. At low pressures, the shorter the discharge pulse (less contamination from the walls), the lower the coefficient. The criterion for electron-ion recombination used by all of these workers was the linear relation between the time and the reciprocal of the electron density. Characteristic of all of these cases also was the fact that the time range within which the recombination process was measured was less than the fundamental diffusion decay time. All of these facts, coupled with the results of the present investigation, support the conclusion that the previously measured recombination coefficients in hydrogen were not values characteristic of the pure gas.

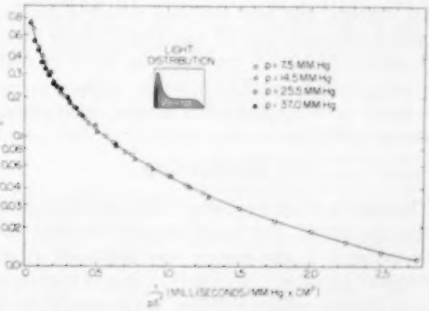


FIG. 4. The data of Fig. 3 normalized to the same initial electron densities.

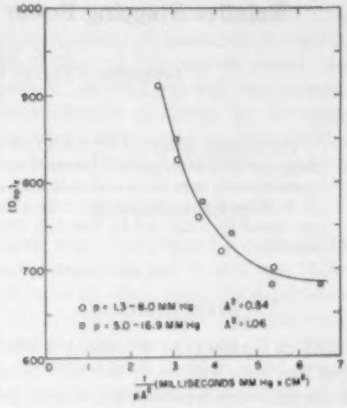


FIG. 5. Variation of the diffusion coefficient with time in the afterglow.

Relative Stopping Power of Various Metals for 20-Mev Protons*

C. P. SONETTI† AND K. R. MACKENZIE
Department of Physics, University of California, Los Angeles, California
(Received May 9, 1955)

The stopping powers of Ni, Cu, Cb, Pd, Ag, Cd, In, Ta, Pt, Au, and Th, relative to Al were measured using the 20.6-Mev external proton beam of the University of California at Los Angeles cyclotron. The measurements were made with sufficient precision so that a graph of $\ln Z$ *vs* stopping power/electron relative to Al showed some fluctuations from a linear relationship. For Th the value differs by 2.6% in the direction found by Kelly and by Teasdale. The results are compared with the data compiled by Lindhard and Scharff.

I. INTRODUCTION

INTEREST in the stopping of protons of intermediate energies has been limited in the past by the complexity of the processes involved, and also by the feeling that the basic mechanisms were understood. Only recently[1] have attempts been made to re-examine the validity of the Fermi-Thomas field approximation used by Bethe[2] and Bloch[3] in their quantum mechanical derivations of a stopping equation.

They used a smooth field approximation because of the impossibility of evaluating an equation containing all likely transition probabilities which would occur in the scattering of atomic electrons by an incident charged particle.

However it might reasonably be expected that the smoothly varying value of stopping power per electron would show fluctuations in value as one examined neighboring elements. In particular such an effect might be expected to be enhanced whenever the wave functions for outer electrons change radically, as when passing from a transition element such as Ni to the normal electronic configuration of Cu. Changes in wave functions will also occur as the phase of the stopping substance is changed. For example, one would expect the stopping power of gaseous Ni to be different than that for Ni in the solid state, the band structure modifying the wave functions and consequently the transition probabilities.

It also might be expected that since transition probabilities also depend upon level densities, the effect of changes in outer electronic configuration would be many times more important than one would estimate on the basis of an arithmetic average over all electrons in an atom; i.e., outer configuration changes would affect stopping power per electron disproportionately.

In view of these expected variations it was planned to measure the stopping power of adjacent foils, insofar as practical, and to make the measurements with the utmost care. This latter was essential, since all previous experiments of this type have shown that such fluctuations would necessarily be less than one percent, where fluctuation will henceforth be defined as the difference between the actual value of stopping power and that found from an arithmetic fit to all the data obtained in this experiment. It should be noted that the results are relatively insensitive to the mode of description since differences between adjacent elements are of prime interest.

II. METHODS AND INSTRUMENTATION

The source of charged particles for this experiment was the external proton beam of the UCLA synchrocyclotron. The beam energy was 20.6 Mev with a spread of ± 0.2 Mev at half maximum. The intensity was approximately 10^{-10} ampere as measured through a $\frac{1}{4}$-in. diameter collimator placed 12 feet from the magnet.

Since fluctuations in stopping power can be measured as accurately in a relative stopping power experiment as in an absolute experiment, it was necessary only to compare the test metals to some standard, which in the case of this experiment was Al. The cyclotron beam was first passed through a collimator and then through the test foil, after which it entered a double Faraday cup which is described below. When balance had been achieved so that each half of the Faraday cup received equal current, the test foil was removed and an Al standard inserted. The Al was then rotated about an axis perpendicular to the beam so as to increase its effective thickness until its stopping power was equivalent to the previous test foil. The equipment placement is shown in Fig. 1.

Faraday Cup

The Faraday cup was composed of two collectors, one inside the other. The beam entered the first (outer) collector and, if the energy was sufficient, it passed through an absorber placed in the cup and then on to the second (inner) collector. When the energy of the beam was properly adjusted the absorber was of precisely the mean range and half the beam was collected on it and therefore in the first collector. This constituted a very sensitive range device when the currents to the two collectors were compared.

* Supported by the joint program of the U. S. Atomic Energy Commission and the Office of Naval Research.
† Now at the Ramo-Wooldridge Corporation, Los Angeles 45, California.
[1] J. Lindhard and M. Scharff, Kgl. Danski Videnskab. Selskab, Mat-fys-Medd. **27**, 15 (1953).
[2] H. Bethe, Ann. Physik **5**, 325 (1930).
[3] F. Bloch, Z. Physik **81**, 363 (1933).

A magnetic field was placed across the Faraday cup so as to eliminate electron emission from the entrance window, and charge exchange between the two halves of the Faraday cup. Comprehensive tests were made which showed that the effects of secondary electrons and ionization currents were negligible.

During the course of the experiment, extra absorbers were always placed in the beam between the foil being tested and the Faraday cup. These were of such a thickness that the straggling energy spectrum was always centered at the rear edge of the Faraday cup absorber.

A difference amplifier was plugged into the end of the Faraday cup and the amplified signal was then brought out to the control room where it was read on a galvanometer. The sensitivity of this system was such that a change of about 5μ g/cm^2 of Al absorber could be detected, although in practice a limit of about 20μ g/cm^2 equivalent was used, this being a limit imposed by cylotron instability.

The amplifier used was not stabilized by feedback, except in the foil uniformity tests discussed later. It was found that in a region of small change of beam intensity the effect of tube variations was negligible and so changing the total beam current by small amounts within any one comparison had no appreciable effect.

A Teflon dielectric condenser shimmed to the correct capacity was constructed on the inner rear wall of the Faraday cup housing. This capacitor was connected to the inner collector and made its capacity the same as that of the outer collector, which equalized time constants thus eliminating transients in the amplifier due to sudden beam intensity variations. Figure 1 shows a cross section of the Faraday cup.

Foil Comparator

It was highly desirable to be able to rapidly and repeatedly compare any test foil with its standard. This was accomplished by mounting a sliding frame in steel ways, the frame containing both a test foil and its standard. The frame was driven by a pneumatic piston actuated from the cyclotron control room, allowing either of the foils to be placed in the beam. Since it was impossible to make up Al standards of stopping power exactly equivalent to the metals tested, they were always made of less thickness and were turned so as to increase their effective thickness. By this means exact relative stopping power could be found. The angle of turning was also set from the control room by means of selsyns.

TABLE I. The surface densities and stopping powers relative to aluminum.

Metal	Surface density mg/cm^2 Group A	Group B	Mass stopping power relative to aluminum (average of groups A and B)	Stopping power per electron relative to aluminum (average of groups A and B)
Ni	55.72	56.13	0.859±0.004	0.868±0.004
Cu	26.75	27.56	0.836±0.006	0.883±0.006
Cb	32.26	32.05	0.766±0.004	0.837±0.004
Pd	60.20	60.82	0.720±0.006	0.805±0.006
Ag	54.32	54.71	0.717±0.006	0.793±0.006
Cd	19.77	19.56	0.714±0.004	0.806±0.004
In	50.46	54.62	0.707±0.010	0.798±0.010
Ta	86.3	91.98	0.607±0.006	0.725±0.006
Pt	117.8	117.2	0.590±0.012	0.712±0.012
Au	95.95	94.65	0.588±0.006	0.708±0.006
Th	67.42	···	0.581±0.006	0.723±0.006

Foil Preparation

Each metal used was assayed spectrographically to determine the impurities and in each case these were so small as to have an affect upon the experiment of less than 0.1%. Small triangles approximately three square inches in area were cut out of a sheet of each metal. These were washed in ether, dried, weighed, and the areas found by measuring the lengths of the sides with an optical comparator to ±0.2 mil. The surface densities found for two complete sets of foils varied from 19 to 117 mgm/cm^2 and are listed in Table I. The measurements were repeated as a check after the experiment was completed. The same process was applied to a large number of Al foils of $\frac{1}{2}$ mil and 1 mil thickness and these were then made up into standards, each test metal having its own Al standard. In a few cases the test foil thickness were such that the energy losses could be directly compared. These served as convenient cross checks.

The accuracy of the experiment is dependent on the uniformity of the foils which was measured in the following way. Each foil was placed in the beam (collimated to $\frac{1}{4}$-in. diameter) and scanned for variations in thickness. This was done by mounting the foil in a ring which rotated with a three-second period and could be offset by different amounts from the beam center. A polar scan resulted and an estimate of variation in thickness was thus found.

The aluminum standards showed no measurable variations when scanned. They were also compared to each other in the beam in order to check the determination of surface densities found by weighing and measurement of areas. These values of relative stopping power agreed in general to better than 0.1% with the directly determined values.

Many elements were omitted from the experiment because the nonuniformity was such as to render the data of no interest.

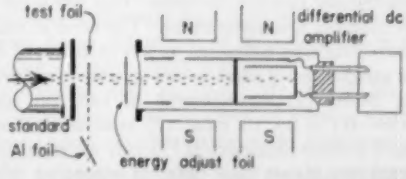


FIG. 1. Equipment arrangement showing differential Faraday cup. Potentials could be applied to the electrode ahead of the first cup to investigate the effect of residual gas ionization, secondary emission, etc.

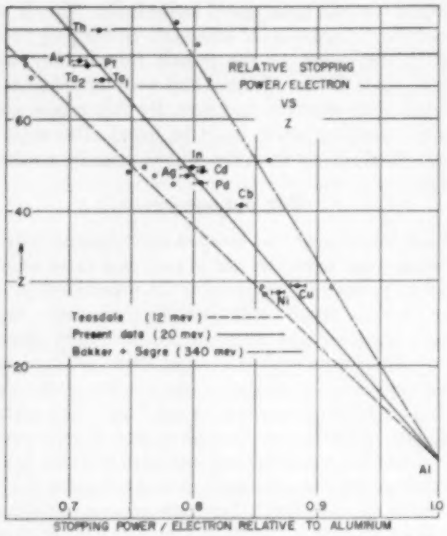


FIG. 2. Stopping power per electron as a function of $\ln Z$.

III. RESULTS AND DISCUSSION OF ERRORS

After all the comparisons had been made, the results were corrected for a multiple scattering range correction[4] as well as being brought to constant energy, the energy loss being different for the various metals tested. The relative values of stopping power per electron were computed from these corrected values of relative mass stopping power.

A further correction was necessary because the straggling spectrum at the rear edge of the absorber in the first Faraday cup changed by a small amount when a test metal was replaced by aluminum. This correction became significant only for high Z values In the case of the Th-Al comparisons, a change in the width of the straggling spectrum of 1.5% at half-maximum would be expected. If the comparisons could have been made with the currents exactly balanced in the two halves of the Faraday cup, there would have been no correction. However, because of fluctuating cyclotron conditions, a state of slight unbalance usually existed. This lack of balance was never more than one-third of the straggling width and so the correction was never large. Since the correction was sensitive to intensity, it was necessary to continuously monitor the beam. This was done with an ion chamber and electrometer. The error in making this correction was estimated as not in excess of 0.1%

The assignment of other errors was made on the following basis: (1) measurement of test foil surface densities—0.1% for all foils; (2) measurement of Al standard surface densities—0.1%; (3) an error arising as mentioned above from different straggling widths, and caused by inability to always operate at the balance point, or mean range—0.1% maximum; (4) measurement of the angle through which the standard foil was rotated (the angle was measured on each side of zero)—0.1%; (5) measurement of foil uniformity was the most uncertain quantity. Nothing was noted in the scanning procedure except a simple periodicity (in one case a variation of 6%) which was presumably caused by a wedge-shapedness of the foils, a characteristic which is common to foil which is formed by a rolling process. The wedge shape might have been ignored to a first approximation as the measured values of surface density were based upon approximately equilateral triangles whose centers, in the relative stopping power experiment lay on the axis of the beam. All the errors were added rms-wise and are indicated in Table I. Since the errors are of the order of $\frac{3}{4}$% it follows that the major fraction must be attributed to foil nonuniformities

In Fig. 2, the values of the stopping power per electron relative to aluminum are plotted *versus* $\ln Z$ and deviations from a straight line are seen. These take the form of fluctuations rather than a smooth departure. The most noticeable deviations are for Ni, Cu, and Th. In the case of Ni and Cu there is a temptation to associate the deviation, if real, with the marked difference in the outer electronic configurations of Ni and Cu.

In the case of the other fluctuations, it is felt that they are not large enough so that anything but a trend may be spoken of. For Th the deviation is marked and agrees in sign with that previously reported by Kelly[5] and by Teasdale.[6] Thorium, being a transition element, could be expected to display an anomalous stopping power. However, its structure is very complex and it might be impossible to predict its stopping behavior.

It is interesting to note that soft x-ray data (emission spectra of band electrons[7]) does not agree with the present measurement which shows that the stopping power per electron is less in Ni than in Cu. This is based on a very crude estimate of the mean ionization potential for the band structures of these elements, and as Ni appears to have a lower mean potential, it would necessarily have a larger stopping power.‡

Lindhard and Scharff[1] have devised a method of plotting stopping powers which allows data taken at different energies with different incident particles to be directly compared.[8] They have collected most of the earlier data on a graph which is reproduced in Fig. 3.

To this graph have been added the data of Kelly (37-Mev alpha particles), Teasdale (12-Mev protons), the data of the present experiment, and the results of Brolley and Ribe.[9] The first three experiments measured

[4] R. Mather and E. Segrè, Phys. Rev. **84**, 191 (1951).

[5] E. L. Kelley, Phys. Rev. **75**, 1006 (1949).

[6] J. G. Teasdale, Office of Naval Research Technical Report No. 3, University of California at Los Angeles, 1949 (unpublished).

[7] F. Seitz, *The Modern Theory of Solids* (McGraw-Hill Book Company, New York, 1940).

‡ *Note added in proof.*—Recent measurements on other elements in a transition region, by a different method, very tentatively suggest variations, possibly due to outer electron configuration and solid state structure, and also suggest a very much smaller Cu Ni anomaly.

[8] Described also in review article by S. Allison and S. Warshaw, Revs. Modern Phys. **25**, 779 (1953).

[9] John E. Brolley and F. L. Ribe, private communication [Phys. Rev. **98**, 1112 (1955)].

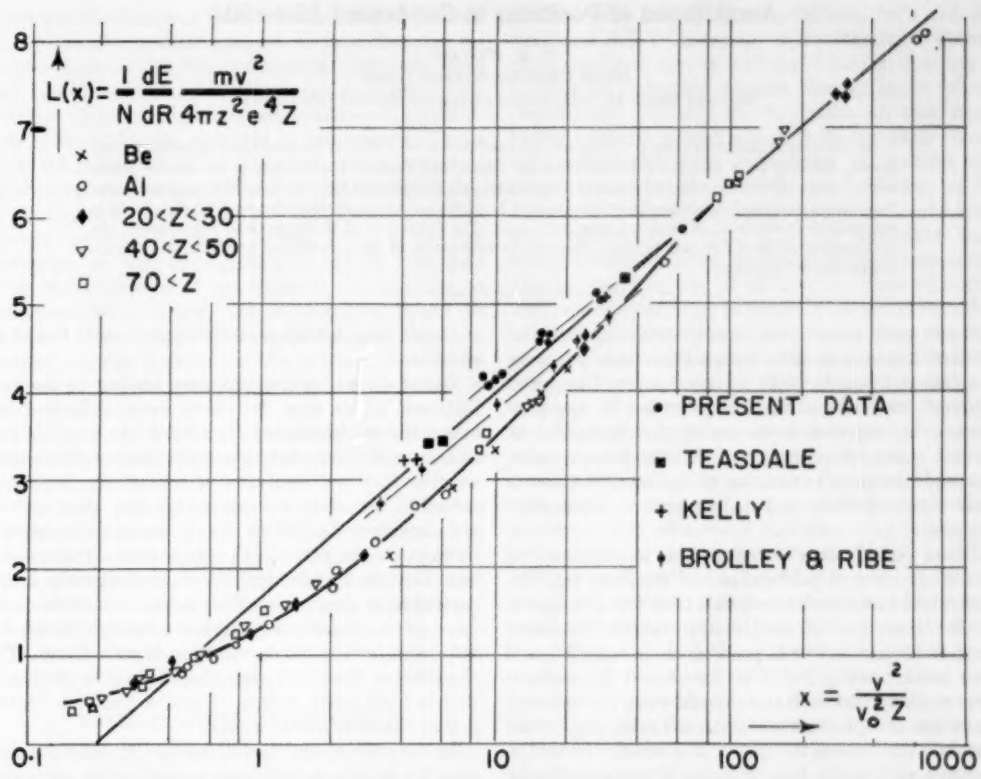


FIG. 3. Data compiled by Lindhard and Scharff, to which have been added present data; points for Al, Au, and Th by Teasdale; and points for Al, Bi, and Th by Kelly. Along a line, which is parallel to a line through the present data, are shown the points for H_2, He, N_2, O_2, Ne, Ar, Kr, and Xe obtained by Brolley and Ribe. ($v_0 = e^2/\hbar$).

relative stopping power, so the point for Al has been arbitrarily put on the straight line which represents the Bloch equation $I = kZ$. The data of Kelly could be fitted, although not very well, to the curved line by placing the Al point on this line, but if this is done for the data of Teasdale, the fit is rather bad. Since the point for Al falls at about the junction of the curved and straight lines for the data of this experiment, there is no possibility of fitting the curve in any other way than shown. There is evidently considerable disagreement with earlier data. The dip in the curve drawn through the earlier experimental values was assumed to be due to a velocity dependent effect. The present experimental results suggest that such an effect is not significant at proton energies over 12 Mev, and that stopping powers can be represented more closely by the Bloch equation.

The absolute stopping power of several gases has been recently measured by Brolley and Ribe[9] using 8.86-Mev deuterons from the Los Alamos cyclotron. The data fit a line of the same slope as the data of the present experiment, but the stopping power per electron for each element is somewhat lower. On the assumption that the high-energy portion of the curve of Lindhard and Scharff is correct on an absolute basis (data of Bakker and Segrè and Mather and Segrè),[10] the separation between these two parallel lines might be interpreted as an indication of a difference between the solid and gaseous state. However, the difference becomes quite small if the data of the present experiment are normalized to the Al point of Sachs and Richardson[11,12] determined with 18-Mev protons in this laboratory. This is the highest point ($x = 54$), to be located on the curve of Fig. 3 by direct measurement of energy loss, and is considered by Sachs and Richardson to be considerably more accurate than their data for the heavier elements where the effects of straggling and multiple scattering are more pronounced.

The authors wish to thank Mr. Steve Plunkett, Mr. George Jones, and members of the cyclotron staff for assistance during the course of the experiment.

[10] C. J. Bakker and E. Segrè, Phys. Rev. 81, 489 (1951); R. Mather and E. Segrè, Phys. Rev. 84, 191 (1951).
[11] Donald C. Sachs and J. Reginald Richardson, Phys. Rev. 83, 834 (1951).
[12] David O. Caldwell, Phys. Rev. 94, 79 (1954).

Annihilation of Positrons in Condensed Materials

P. R. WALLACE
McGill University, Montreal, Canada

(Received March 25, 1955; revised manuscript received July 11, 1955)

An analysis of the experimental results on the annihilation of positrons in condensed materials is suggested, differing from that recently published by Dixon and Trainor. Considerations are offered which make it seem unlikely that excited states of positronium play a significant role. The long lifetime in molecular materials is explained in terms of the formation of $1s$ triplet positronium, while the short lifetime in these materials is interpreted as having a complex origin. The frequency of three-quantum annihilation, the temperature effect of the long lifetime, the angular correlation of the γ rays, and annihilation in superconductors are discussed.

IN the past three years, considerable experimental work has been done on the annihilation of positrons in solids and liquids.[1–8] In a recent paper, Dixon and Trainor[9] have offered an interpretation of the phenomena in question in terms of the formation of excited states of positronium. In the present note, I should like to call attention to serious difficulties in their interpretation, and to suggest an alternative analysis.

Dixon and Trainor's explanation, in terms of an excited $2s$ state of positronium, of the long (τ_2) lifetime found in molecular materials faces two difficulties: (i) the large size of excited-state orbits (the mean particle distance in free $2s$ positronium is ≈ 6.4 A^0), and their weak binding (≈ 1.7 ev for $2s$ and $2p$) make it seem unlikely that such states could exist in condensed materials, and (ii) if excited states *did* exist, they would certainly not possess the spherical symmetry needed to ensure a sufficiently long lifetime. It is shown in the appendix that an initial s-state would very quickly become admixed with p-states, which would then be de-excited to the $1s$ state by collision in a time short compared with the annihilation lifetime. Another possibility is, of course, the radiative decay of the p-state. This has a lifetime $\approx 3\times 10^{-9}$ second, and is therefore less important than de-excitation by collision.

Dixon and Trainor correctly point out, in their appendix, an inconsistency between the relative intensities of the two components and the corresponding lifetimes, on the basis of a model which allows for decay of positrons only in positronium, with the possibility of conversion from triplet to singlet $1s$ state. This difficulty can, however, be removed without recourse to the hypothesis of excited states, if it is admitted that not all positrons decay while bound to electrons.

Before discussing the processes leading to the two lifetimes, let us note the fundamental difference between the substances which exhibit the long lifetime and those which do not. The latter class of substances, whether they are insulators or conductors, is characterized by the presence of valence binding. The electrons are described by Bloch wave functions extending throughout the crystal. The substances in which the long lifetime appears are molecular materials, either crystalline or amorphous. The motion of electrons takes place predominantly only *within* molecules, which are held together by weak van der Waal's forces. The situation in these materials should then be similar to that in inert gases, except, of course, that the density is that characteristic of a solid.

In the case of metals and valence crystals the fact that the lifetime is fairly independent of the particular substance does not necessarily indicate, as assumed by D.T. and by Bell,[10] that the lifetime must be explained only in terms of positronium formation. In fact, as was pointed out in D.T., the lifetime is several times shorter than would be expected of positronium, since only *singlet* positronium has short lifetime and rapid singlet↔triplet exchange must be assumed. It seems likely, therefore, that whether or not positronium is formed, the polarization of the wave functions of *unbound* electrons must make an important contribution to the shortening of the decay lifetime. It is interesting to note that the work of Green and Stewart[8] on the angular correlation of the emitted γ rays is consistent with the picture of annihilation of the positrons by *free* electrons. It is quite conceivable that the net effect of the polarization will be largely determined by conditions in the immediate neighborhood of the positron and therefore be more or less independent of the substance.

In the case of molecular materials, a considerable proportion of positrons should survive the process of successive electron capture and loss and ultimately annihilate when free. Actually, the generally longer and more variable lifetime of the short (τ_1) component in

[1] De Benedetti, Cowan, Konneker, and Primakoff, Phys. Rev. **77**, 205 (1950).
[2] R. E. Bell and R. L. Graham, Phys. Rev. **90**, 644 (1953).
[3] S. de Benedetti and S. J. Richings, Phys. Rev. **85**, 377 (1952).
[4] R. L. Graham and A. T. Stewart, Can. J. Phys. **32**, 678 (1954).
[5] W. E. Millett, Phys. Rev. **94**, 809 (1954).
[6] R. Stump and H. Talley, Phys. Rev. **94**, 809 (1954).
[7] Page, Heinberg, Wallace, and Trout, Phys. Rev. **99**, 665(A) (1955).
[8] R. E. Green and A. T. Stewart, Phys. Rev. **98**, 486 (1955).
[9] W. R. Dixon and L. E. H. Trainor, Phys. Rev. **97**, 733 (1955). Hereafter to be referred to as D.T.
[10] R. E. Bell in Siegbahn, *Beta- and Gamma-Ray Spectroscopy* (Interscience Publishers, Inc., New York, 1954), p. 686.

these materials suggests that the decay is not exclusively that of singlet positronium. So, in fact, does the intensity ratio of the two components. When electrons are captured in $1s$ positronium at sufficiently low energies ($\lesssim$10 ev, say), this positronium will predominantly be elastically scattered in successive collisions with the molecules of the substance, which represent closed-shell configurations. If this positronium is singlet, it will have a lifetime $\approx 1.2\times10^{-10}$ sec. To this contribution to the short life would be added that of the free positrons, as well as positrons chemically combined with the molecules of the material. In addition to being somewhat longer, this contribution should be expected to show some variations from one material to another, though polarization effects would make these variations proportionately less than the variations in electron density.

The experiments are not inconsistent with the view that the short lifetime has a complex origin in molecular materials.

The long (τ_2) component in molecular materials should then be associated with the formation of $1s$ triplet positronium. Its decay is probably due to "pick-off annihilation" as suggested by Dresden[11]; that is, annihilation during collision with an appropriately oriented electron on a target molecule.

The following comments are offered on other particular aspects of the problem:

THREE-QUANTUM ANNIHILATION

As shown by Graham and Stewart,[4] the three-quantum yield follows naturally from the assumption that the τ_3 component is associated with the formation of triplet $1s$ positronium. D.T.'s explanation of the three-quantum counting rate in molecular materials depends on the assumption of a rather rapid triplet→singlet conversion rate in these materials. However, at low energies rapid conversion is not possible. Ore[12] has shown that direct conversion has very long lifetime ($\gtrsim10^{-4}$ sec). The only other possibility is that of electron exchange. However, this would involve a spin-flip of an electron on the target molecule. Since the ground-state is a closed-shell configuration, this can only be accomplished by excitation of the molecule to a higher electron state with spin 1. This will require considerable excitation energy, and so will not take place in slow collisions.

The results on three-quantum annihilation are consistent with the assumption of slow conversion, annihilation taking place by a process such as that of "pick-off" mentioned above.

THERMALIZATION, TEMPERATURE EFFECT, AND ANGULAR CORRELATION OF γ RAYS

With regard to the temperature effect on the τ_2 lifetime observed by Bell and Graham (lengthening of lifetime with increasing temperature), both the discussion of D.T. ("loosening" of $2s$ states due to thermal excitation) and that of Bell and Graham (increase in number of collisions) assume thermalization of the positronium. Although the thermalization time would be very difficult to estimate, it is by no means certain that it is shorter than the lifetime. At energies high enough to excite electronically the molecules of the substance, inelastic scattering with fairly large energy loss ($\approx$volts) may take place. At somewhat lower energies, inelastic collisions involving vibrational excitation are possible, and at all energies there will occur slight energy losses due to elastic collisions. It should be noted that, neglecting electron exchange, these processes must be second-order ones. For the Coulomb forces acting to produce the collision are antisymmetric in exchange of positron and electron, whereas the initial and final wave functions are symmetrical. Scattering takes place therefore only through the agency of intermediate states (e.g., $2p$) of the positronium.

A rough estimate of the probability of inelastic scattering with vibrational excitation may be made as follows: the scattering potential is roughly the variation in the potential of the atoms of the molecule over the zero-point oscillations of the atoms. The amplitude of oscillation of the atoms is $(h/2\omega M)^{\frac{1}{2}}$, ω being the vibrational frequency and M the atomic mass. If the mean gradient of the atomic field is $\bar{\imath}$ ev per A, the mean perturbing potential of a particular atom is then $H'\approx0.05\bar{\imath}/(E_vA)^{\frac{1}{2}}$ ev, A being the nuclear mass of the atom and E_v the vibration energy in volts, the signs depending on whether it acts on electron or positron. It will be noted that this will be much smaller than the potential involved in elastic scattering, which is $\approx\bar{\imath}$.

The cross section will therefore be less than the corresponding one for elastic scattering by a factor $\approx(0.05)^4/E_v^2A^2$. In most cases $E_v\approx0.1$–0.5 ev. With a reasonable assumption for the elastic cross section, this mechanism will make a negligible contribution.

In the case of electron exchange, in each case a first-order scattering is possible. The small amplitude of oscillation of the atoms is then less effective in cutting down the inelastic relative to the elastic cross section [ratio$\approx(0.05)^2/E_vA$]. However, the increase in inelastic cross section due to the presence of a first-order matrix element is at least partially compensated by the fact that this matrix element is of the overlap type and is therefore substantially smaller than that for a direct scattering process.

Consider a case which is particularly favorable for slowing-down—that of ice. The main agency for energy transfer will be collision with the hydrogen atoms. There are about 6×10^{22} of these per cc. The velocity of a positronium atom of energy E ev is $\approx0.4\sqrt{E}\ 10^8$ cm/sec. If we consider a process with a cross section σ, the probability per unit time of that process is $2.4\sqrt{E}\ 10^{30}\sigma$, or the probability that it takes place in 10^{-9} sec

[11] M. Dresden, Phys. Rev. **93**, 1413 (1954).
[12] A. Ore, Naturvitensk. rekke, Universitetet i Bergen **12** (1949).

is $2.4\sqrt{E}\ 10^{21}\sigma$. According to Massey and Mohr,[12] the exchange elastic scattering cross section on hydrogen should be $\pi a_0^2 f = 0.9f\ 10^{-16}$ cm^2 where f is a small number of order unity. Therefore, there will be many elastic scatterings before annihilation. However, inelastic scattering with vibrational excitation should happen less often by a factor $(0.05)^2/E_v$. Possible E_v's for H_2O are ≈ 0.2 ev or ≈ 0.45 ev.

It is evident that this process is of such an order of probability that more careful calculations would be necessary to establish whether or not it would, in 10^{-9} sec, lower the positronium energy below the threshold of E_v in the substances in question. This is particularly the case since the use of Born approximation would overestimate the cross sections, and the case considered (ice) is the most favorable for slowing down.

Below the lowest energy E_v, only elastic collisions with whole molecules are possible. If the molecule has mass A', the mean fractional energy loss per collision will be $4/1840A'$. Reduction of the energy by a factor of 10 (e.g., from 0.25 ev to 0.025 ev) would, assuming constant cross section σ_0, take a time $1840A'/4N\sigma_0 v_0$, v_0 being the thermal velocity and N the number of molecules per unit volume. If $t=10^{-9}$ sec this would mean $N\sigma_0 = 2.9\times 10^5 A'$. For the most favorable case of ice, this implies $\sigma_0 \approx 2\times 10^{-16}$.

Again, it appears that thermalization might just be possible in the favorable case of ice at ordinary temperatures, but it is less likely in substances with heavier molecules, or at lower temperatures.

It seems, therefore, that a discussion of the temperature effect based on the assumption of thermalization of positronium is open to question. An alternative explanation is the following: it may be assumed that positronium moves adiabatically in the nuclear potential, and should thus be found with highest probability in the larger interstices, near the potential minima. Since, with greater molecular motion, some larger molecular separations appear, the positron intensity will be enhanced in these intermolecular "holes" where the intensity of molecular electrons is least, "pick-off annihilation" is less likely and the observed temperature effect might be expected.

In connection with the question of thermalization of positronium in molecular materials, attention should be drawn to the experiments on angular correlation of annihilation γ rays in these materials by Page, Heinberg, Wallace, and Trout,[7] who found a very narrow component, indicating an energy of the mass center of the annihilating pair of one volt or less, which accounted for about 20% of all decays. If in fact positronium is not thermalized, this might be attributed to a component of the *short* lifetime. An attractive hypothesis would be to associate it with the decay of positrons which have, after slowing down, become bound to molecules of the substance.

[12] H. S. W. Massey and C. B. O. Mohr, Proc. Phys. Soc. (London) **67**, 695 (1954).

ANNIHILATION IN SUPERCONDUCTORS

Stump and Talley[4] found a lengthening of the lifetime of positrons in superconducting lead, as compared with ordinary lead. D.T. attribute this to the tendency of superconducting electrons not to exchange energy with impurities, so that $2s$ positronium, once formed, can be maintained. However, it is difficult to see why, if superelectrons do not react with positronium, they should not be indifferent to slow positrons as well, with consequent inhibition of positronium formation. In fact, such a conclusion is also suggested by the supposed low effective mass of superelectrons. This failure to attract electrons would itself give rise to a lengthening of the lifetime.

APPENDIX

We shall first assume the possibility of $2s$ positronium in a molecular solid, and estimate roughly the probability that it is transformed to the $2p$ state during scattering. Using Born approximation, this probability is

$$\frac{1}{\tau} = \frac{2\pi}{\hbar}|\gamma(\mathbf{k},\mathbf{k}')|^2_{E'=E} d\Omega(\mathbf{k}')\rho(E). \tag{1}$$

$\rho(E)$ is the density of final states per unit solid angle and the integration is over final directions.

$$\gamma(\mathbf{k},\mathbf{k}') = \int e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}}\varphi_{2s}(\boldsymbol{\varrho})\varphi_{2p_n}(\boldsymbol{\varrho}) \times[V(\mathbf{R}+\tfrac{1}{2}\boldsymbol{\varrho})-V(\mathbf{R}-\tfrac{1}{2}\boldsymbol{\varrho})]d^3\mathbf{R}d^3\boldsymbol{\varrho}. \tag{2}$$

Here $\mathbf{k}, \mathbf{k}'$ are the initial and final wave-number vectors of the positronium, φ_{2s} and φ_{2p_n} the positronium wave functions for the $2s$ state and the $2p$ state with axis $\mathbf{n}$, respectively; $\mathbf{R}$ is the vector specifying the center of mass of the positronium and $\boldsymbol{\varrho}$ the position vector of electron relative to positron. The initial and final wave functions are normalized to unit volume, and V is the potential of the electron in the Coulomb field of a scattering molecule. For a rough approximation the effect of the Pauli principle on the electron is ignored.

If $W(\mathbf{K})$ is the Fourier transform of $V(\mathbf{r})$, this integral is easily shown to be

$$\gamma(\mathbf{k},\mathbf{k}') = iW(\mathbf{K})\frac{6K^3a_0^3}{(K^2a_0^2+1)^4}\cos(\mathbf{n},\mathbf{K}), \tag{3}$$

where a_0 is the Bohr radius, and $(\mathbf{n},\mathbf{K})$ is the angle between the axis of the p-state wave function and the vector $\mathbf{K}=\mathbf{k}-\mathbf{k}'$.

Summing $|\gamma|^2$ over all possible orientations of the p-state wave function, the cosine term will drop out. We may also average over all orientations of the

molecule, i.e., replace $|W(\mathbf{K})|^2$ by

$$I=\frac{1}{4\pi}\int |W(\mathbf{K})|^2 d\Omega_{\mathbf{K}}. \tag{4}$$

Using the Fourier integral expression for $W(\mathbf{K})$ this yields

$$I=\int V(\mathbf{r})V(\mathbf{r}')\frac{\sin K|\mathbf{r}-\mathbf{r}'|}{K|\mathbf{r}-\mathbf{r}'|}d^3\mathbf{r}d^3\mathbf{r}'. \tag{5}$$

We may, for purposes of rough estimation, ignore contributions from $K|r-r'|>\pi/2$. Because this will involve an underestimate, we can replace the $\sin K|\mathbf{r}-\mathbf{r}'|/K|\mathbf{r}-\mathbf{r}'|$ by unity within these limits. But $|6K^3a_0^3/(K^2a_0^2+1)^4|^2$ contributes significantly only in the range $0.4\leq Ka_0\leq 1.4$. Thus, if $|\mathbf{r}-\mathbf{r}'|<(\pi/2)(a_0/1.4)$, over the whole significant range of K we may write approximately

$$I=\int V(\mathbf{r})V(\mathbf{r}')d^3\mathbf{r}d^3\mathbf{r}',$$

the integral being taken over the limits of $|\mathbf{r}-\mathbf{r}'|$ indicated. We shall neglect contributions to I from larger $|\mathbf{r}-\mathbf{r}'|$. Then, without specifying the details of $V(\mathbf{r})$ we see that

$$I\approx \bar{V}^2 V_M \pi a_0^3,$$

where $\bar{V}$ is the mean molecular potential, V_M the volume of the molecule.

To obtain the total probability per unit time $1/\tau_0$ for scattering from a $2s$ to a $2p$ state we then multiply (1) by $1/V_M$, the number of molecules per unit volume, and get

$$\frac{1}{\tau_0}=\frac{2\pi}{\hbar}\bar{V}^2\pi a_0^3\frac{1}{4\pi^3}\frac{mk}{\hbar^2}\int\frac{36K^6a_0^6}{(K^2a_0^2+1)^6}d\Omega. \tag{6}$$

To do the integral over direction, we use the fact that

$$\begin{aligned}K^2a_0^2&=2k^2a_0^2(1-\cos\theta)\\&=2(E/W_0)(1-\cos\theta)=2\eta(1-\cos\theta).\end{aligned}$$

W_0 being the energy of the $1s$ state of positronium. For $\eta\approx 1$, the integral in (6) is approximately $9\pi/35\eta$. Thus, finally,

$$\frac{1}{\tau_0}=\frac{9}{280\eta^4}\frac{\bar{V}}{W_0}\frac{\bar{V}}{\hbar}. \tag{7}$$

This gives a time of the order of 10^{-14}–10^{-15} sec.

Consider next scattering with de-excitation from a $2p$ state to the $1s$ state. The matrix element is similar to that of (2), except that $\varphi_{2s}(\rho)$ is replaced by $\varphi_{1s}(\rho)$, and in addition $|\mathbf{k}'|$ is no longer equal to $|\mathbf{k}|$ since energy is gained on collision. On evaluating we find

$$\gamma'(\mathbf{k},\mathbf{k}')=iW(\mathbf{K})\frac{24\cos\alpha}{\sqrt{2}}\frac{Ka_0}{(K^2a_0^2+8/3)^3}, \tag{8}$$

α being the angle between $\mathbf{K}$ and the axis of the $2p$ wave function. Averaging $|\gamma'|^2$ over all possible directions for this axis, we can replace $\cos\alpha$ by $1/\sqrt{3}$. This time $k'^2=k^2+(4m/\hbar^2)\frac{3}{4}W_0$ or $k'^2a_0^2=k^2a_0^2+\frac{3}{4}$, from which it follows that

$$K^2a_0^2=2\eta+\tfrac{3}{4}-2[\eta(\eta+\tfrac{3}{4})]^{\frac{1}{2}}\cos\theta. \tag{9}$$

The range of Ka_0 over which $(Ka_0)^2/(K^2a_0^2+8/3)^6$ has significant values is comparable to that of the previous calculation; by similar arguments therefore

$$\frac{1}{\tau'}=\frac{2\pi}{\hbar}\bar{V}^2\pi a_0^3\frac{1}{4\pi^3}\frac{mk'}{\hbar^2}\int\frac{96K^2a_0^2}{(K^2a_0^2+8/3)^6}d\Omega \tag{10}$$

approximately. The integral can be evaluated, and yields

$$\frac{8\pi}{5[\eta(\eta+\frac{3}{4})]^{\frac{1}{2}}}\left\{\frac{(77/4)+30\eta-30[\eta(\eta+\frac{3}{4})]^{\frac{1}{2}}}{[(41/12)+2\eta-2(\eta(\eta+\frac{3}{4}))^{\frac{1}{2}}]^5}-\frac{(77/4)+30\eta+30[\eta(\eta+\frac{3}{4})]^{\frac{1}{2}}}{[(41/12)+2\eta+2(\eta(\eta+\frac{3}{4}))^{\frac{1}{2}}]^5}\right\}. \tag{11}$$

This expression varies with η, though not too rapidly in the interesting energy range ($\eta\approx 1$). Writing $\frac{1}{5}$ of the curly bracket as $F(\eta)$

$$\frac{1}{\tau'}=\frac{1}{\sqrt{\eta}}\frac{\bar{V}}{W_0}\frac{\bar{V}}{\hbar}F(\eta). \tag{12}$$

At $\eta=1$, $F(\eta)=0.0114$, so that the mean time for this transition at this energy is about three times that for $2s\rightarrow 2p$ conversion, and is therefore still very short compared with the lifetime.

ACKNOWLEDGMENTS

The author is indebted to the Summer Research Institute of the Canadian Mathematical Congress for a summer fellowship during the tenure of which this work was carried on.

Magnitude of the Splitting of the Lowest Level in Chromium Alum*

PAUL H. E. MEIJER, *Duke University, Durham, North Carolina*

AND

H. J. GERRITSEN, *Kamerlingh Onnes Laboratorium, Leiden, Netherlands*

(Received May 31, 1955)

This note contains the results of a recalculation of the lowest energy levels of the chromium ion in chromium potassium alum. After diagonalization of the Hamiltonian of the electric field, the spin-orbit coupling is introduced and the spin Hamiltonian of the lowest level calculated by a formal method. The numerical value differs from those published previously. The result is a different value for the parameter a (radial average of the cylindrical field).

THE energy difference between the two doubly degenerate states of the lowest level in chromium alum can be expressed as: $\delta = c\Delta\lambda^2/(h\nu_c)^2$, where λ is the spin orbit coupling constant; $h\nu_c = 15\bar{Q}$ the distance between the levels Γ_3 and Γ_5 due to the cubical field, and $\Delta = (15/2)\bar{a}$ the strength of the cylindrical field expressed in the splitting of Γ_5, due to this field. The parameters $\bar{a}$ and $\bar{Q}$ are the radial averages of the coefficients a and Q in the Hamiltonian of the electrical field:

$$\mathfrak{H}_{el} = Q(r)\{(10)^{\frac{1}{2}}(Y_4^2 - Y_4^{-2}) + (7)^{\frac{1}{2}}Y_4^0\} + a(r)Y_2^0. \quad (1)$$

The numerical constant c has been found to be 152/9 in our calculation, about twice Van Vleck's[1] value.[2] This has some influence upon the results quoted by Weiss.[3] The value of $\bar{Q}$ is practically unchanged, but $|\bar{a}|$ becomes 82 cm^{-1}. Our result also disagrees with Broer's calculation[4] which gave $c = 88/9$.

Calculation.—The expressions of Becquerel and Opechowski[5] have the advantage that the influence of the cylindrical field is already included, so we do not have to perform a perturbation calculation of the third order. We will label their transformed orbital wave functions by $|i\rangle$ $(i = 1, \cdots, 7)$, and avoid the somewhat awkward coefficients by a series development: $(\bar{a}/\bar{Q} = A)$

$$\alpha_2 = (5/2)^{\frac{1}{2}}(1 + A/3), \qquad \alpha_4 = -4(1 - A/3)/(10)^{\frac{1}{2}},$$
$$\beta_4 = -(1/5)^{\frac{1}{2}}(1 - A/4), \qquad \beta_5 = 5^{\frac{1}{2}}(1 + A/4),$$
$$M_2 = (2/9)^{\frac{1}{2}}(1 - 5A/27), \qquad N_5 = (1/6)^{\frac{1}{2}}(1 - 5A/24).$$

Multiplying $|i\rangle$ by the spin functions, we now have to solve the equation:

$$(\mathfrak{H}_{el} + \mathfrak{H}_{so})U = \sum_j \sum_{m'} [E_i^0\langle m|\delta|m'\rangle + \langle im|\mathfrak{H}_{so}|jm'\rangle]U_{jm'} = EU. \quad (2)$$

The degeneracy of the level i is not lifted in the first approximation, so we make a transformation of the state vector $U = (\exp S)V$ resulting in the Hamiltonian:

$$e^{-S}\mathfrak{H}e^{S} = \mathfrak{H}_{el} + \mathfrak{H}_{so} + [\mathfrak{H}_{el}, S] + [\mathfrak{H}_{so}, S] + \tfrac{1}{2}[[\mathfrak{H}_{el}, S], S] + \cdots. \quad (3)$$

The condition to remove the first-order perturbation $\mathfrak{H}_{so} + [\mathfrak{H}_{el}, S] = 0$ determines S. Substitution in the second-order terms in (3) gives:

$$\mathfrak{H}_{so} + \tfrac{1}{2}[[\mathfrak{H}_{el}, S], S] = \tfrac{1}{2}[\mathfrak{H}_{so}, S].$$

The diagonal element with regard to the lowest i-level is

$$\tfrac{1}{2}\langle 7|[\mathfrak{H}_{so}, S]|7\rangle = \langle 7|\mathfrak{H}_{so}|i\rangle\langle i|\mathfrak{H}_{so}|7\rangle/(E_i - E_7).$$

In our case this element happens to be also diagonal with respect to m [the potential (1) can give rise to off-diagonal elements in the spin matrix only if $S \geqslant 2$] so the expression represents already the second-order correction to the energy:

$$E^{(2)} = \lambda^2 \sum_k \sum_l \sum_i S_k S_l \langle 7|L_k|i\rangle\langle i|L_l|7\rangle/(E_i - E_7) \quad (4)$$

$(l, k = +, z, -)$, where $S_\pm = 2^{-\frac{1}{2}}(S_x \pm iS_y)$, etc. The numerator is exactly zero for $i = 3$ and zero up to first order terms for $i = 1, 2$. In the computation the only relevant elements are: $\langle 4|L_-|7\rangle$, $\langle 5|L_+|7\rangle$ and $\langle 6|L_z|7\rangle$, with the result:

$$E^{(2)} = 18\lambda^2 M_2^2[(1 + 5A/9)(S_+S_- + S_-S_+)/(-15\bar{Q} + 5\bar{a}/2) + S_z^2/(-15\bar{Q} - 5\bar{a})],$$

from which we derive

$$\delta = E_{\pm\frac{1}{2}} - E_{\pm\frac{3}{2}} = -\frac{\lambda^2\bar{a}}{(15\bar{Q})^2}\,\frac{380}{3}.$$

As the resonance experiments show that the $m = \pm\frac{3}{2}$ level is lowest, $\bar{a}$ has to be negative.

* A part of this work was supported by the U. S. Air Force through the Office of Scientific Research of the Air Research and Development Command.

[1] J. H. Van Vleck, J. Chem. Phys. **7**, 61 (1939).

[2] Professor Van Vleck was very kind to confirm that there was an error in his formula. He had performed a third-order calculation in which the terms

$$\langle\Gamma_2|\mathfrak{H}_{el1}|\Gamma_5\rangle\langle\Gamma_5|\mathfrak{H}_{so}|\Gamma_4\rangle\langle\Gamma_4|\mathfrak{H}_{so}|\Gamma_2\rangle$$

were neglected. $\mathfrak{H}_{el1}$ stands for the second term in (1) which is here a part of the perturbation.

[3] P. R. Weiss, Phys. Rev. **73**, 470 (1948).

[4] L. J. F. Broer, Physica **9**, 547 (1942).

[5] J. Becquerel and W. Opechowski, Physica **6**, 1039 (1939).

Liquid Helium Vapor Pressure Equation

J. R. Clement, J. K. Logan, and J. Gaffney
United States Naval Research Laboratory, Washington, D. C.

(Received May 27, 1955)

An empirical equation for the vapor pressure of liquid helium has been developed. This equation is valid throughout the entire range of temperatures below the critical point, and is in accord with the latest available data on the vapor pressure to within about ±0.002°K.

THE problem of representing the vapor pressure of liquid helium by analytic functions has been solved in the past by means of three empirical equations, one valid in the temperature range below the λ point (2.17°K), a second valid between the λ point and the normal boiling point (4.22°K), and still a third between the boiling point and the critical point (5.20°K).[1] In 1948 a table, prepared at the Kamerlingh Onnes Laboratory in Leiden, giving the vapor pressure as a function of temperature was adopted as the definition of an "agreed" scale which all cryogenic laboratories should use in reporting data.[2] The accuracy of the data in this table had already been questioned some two years prior to its adoption.[3,4] Extensive research during the past two years[5–10] has demonstrated beyond question that there exist small, but significant, differences between the 1948 scale and the true thermodynamic scale (~0.01° at 2.2°K and ~0.02° at 4.9°K). A recent examination of the 1948 scale by the authors[11] further verified these major errors and, in addition, revealed point-to-point roughness in the scale itself. In particular, some rather large and abrupt changes in slope of the 1948 scale were found, the most notable of these occurring near the λ point, near 2.8°K and near 4.2°K.

During the course of the authors' investigation of the 1948 scale, the derivative of the logarithm of the vapor pressure with respect to reciprocal temperature ($d\ln P/dT^{-1}$) was calculated, directly from the 1948 scale at first and subsequently from thermodynamic data. These data were found to fit the following empirical function rather well over the entire temperature range below the critical point,

$$-d\ln P/dT^{-1}=A+BT+CT^2-D\tan^{-1}(\alpha T-\beta), \quad (1)$$

where A, B, C, D, α, and β are adjustable constants. Integration of this equation yields

$$\ln P=I-\frac{A}{T}+B\ln T+\frac{C}{2}T^2 -D\left(\frac{\alpha\beta}{\beta^2+1}-\frac{1}{T}\right)\tan^{-1}(\alpha T-\beta) -\frac{\alpha D}{2(\beta^2+1)}\ln\frac{T^2}{1+(\alpha T-\beta)^2}, \quad (2)$$

where I is the constant of integration. The constants in Eq. (2) were determined from the 1948 scale plus an average of all measured corrections with the following results: $I=4.6202_5$, $A=6.399$, $B=2.541$, $C=0.00612$, $D=0.5197$, $\alpha=7.00$, and $\beta=14.14$. Pressure, P, is in mm Hg (20°C and standard gravity) and $-\pi/2<\tan^{-1}(\alpha T-\beta)<\pi/2$ radians.

Figure 1 demonstrates graphically the results. The various experimental data points plotted in Fig. 1 are, with the exception of the thermodynamic calculation of Erickson[12] and the isotherms of Keller, an average of several individual measurements, all at temperatures within about 0.1°, or less, of the averaged temperature at which the point is plotted. The rather jagged curve drawn in the figure was obtained in the following manner. The pressure corresponding to each 0.01° temperature interval between 0.66°K and 5.20°K was calculated from Eq. (2).[13] At each of these calculated pressures, the corresponding value of T_{48} was found from the "agreed" scale. The difference between T_{48} and the temperature used in the calculation, T, was plotted at each value of T and the individual points were connected by straight lines.

Equation (2) appears to eliminate most of the difficulties previously encountered in consolidating measure-

[1] W. H. Keesom, *Helium* (Elsevier Publishing Company, Inc., Amsterdam, 1942), Chap. IV.
[2] H. van Dijk and D. Shoenberg, Nature **164**, 151 (1949).
[3] J. Kistemaker and W. H. Keesom, Physica **12**, 227 (1946).
[4] J. Kistemaker, Physica **12**, 272 (1946).
[5] R. A. Erickson and L. D. Roberts, Phys. Rev. **93**, 957 (1954).
[6] Worley, Zemansky, and Boorse, Phys. Rev. **93**, 45 (1954).
[7] R. Berman and C. A. Swenson, Phys. Rev. **95**, 311 (1954).
[8] W. E. Keller, Phys. Rev. **97**, 1 (1955).
[9] W. E. Keller, Phys. Rev. **98**, 1571 (1955). (The authors are indebted to Dr. Keller for an advance copy of his paper.)
[10] Corak, Garfunkel, Satterthwaite, and Wexler, Westinghouse Research Report 60-94466-5-R1, December, 1954 (unpublished).
[11] Clement, Logan, and Gaffney, Naval Research Laboratory Report No. 4542, May, 1955 (unpublished).
[12] R. A. Erickson (private communication). A re-calculation of the helium vapor pressure scale from thermodynamic data as done by Erickson and Roberts in reference 5, using the virial coefficient data reported by Kilpatrick, Keller, and Hammel [Phys. Rev. **97**, 9 (1955)]. (The authors are indebted to Dr. Erickson for a tabulation of the numerical results of his calculations.)
[13] Values of vapor pressure as given by Eq. (2) have been calculated by the Naval Research Laboratory electronic digital computer (NAREC) at each 0.005° temperature interval between 0.6° and 5.3°K, thanks to the kind assistance and cooperation of Dr. B. Lepson of this laboratory. Copies of these calculations may be had by writing to the authors.

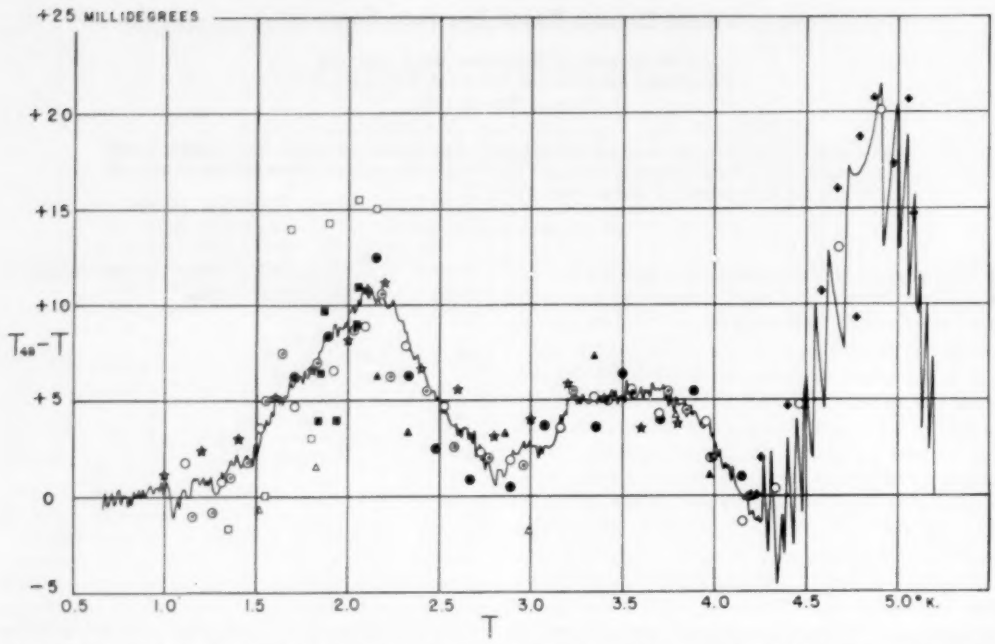


FIG. 1. Summary of data on errors in the 1948 liquid helium vapor pressure scale. The jagged curve is the difference between temperature as defined by the 1948 scale and by Eq. (2) of this report. Plotted points are measurements of ■ Kistemaker and Keesom,[3] □ Kistemaker,[4] ⊙ Erickson and Roberts (magnetic thermometer data of reference 5), + Berman and Swenson,[7] ▲ Keller,[8] △ Keller[9], ○ Corak, *et al.*,[10] ● Clement, *et al.*,[11] and ★ Erickson.[12]

ments of the vapor pressure of liquid helium. First, it is a single equation valid over the entire temperature range below the critical point. Second, it can be calculated to any precision desired, eliminating entirely the possibility of any point-to-point roughness or abrupt change in slope as found in the 1948 scale. Third, it seems to represent the relationship between the vapor pressure and the true thermodynamic temperature within $\pm 0.002°$ or less, a deviation comparable in magnitude to the uncertainty in our present knowledge of this relationship.

Note added in proof.—Since the preparation of this report, a number of new results have appeared which lead to some modification of the conclusions stated above. First, Keller has made some corrections in the analysis of his He^4 isotherm data (reference 8). Second, the data of Berman and Swenson (reference 7) have been re-examined in light of a redetermination of the boiling point of normal hydrogen reported by Moessen, Aston, and Ascah [J. Chem. Phys. **22**, 2096 (1954)]. Third, new magnetic measurements (reported at the International Conference on Low Temperature Physics, Paris, September 2–8, 1955) have been made by Ambler and Hudson at the National Bureau of Standards. These new and corrected data show systematic deviations from Eq. (2), the maximum deviation being about 5 millidegrees. The carbon thermometer calibration data of Corak *et al.* [reference 10, an abridged version of which appeared in Phys. Rev. **98**, 1699 (1955)] are consistent with these deviations.

During the recent Conference on Low Temperature Physics in Paris the vapor pressure-temperature relationship for He^4 was discussed and replacement of the 1948 scale was recommended. Part of the recommendation was that the conversion of vapor pressure to temperature, when the vapor pressure is measured in a bulb, be based on the experimental results mentioned in this note. A vapor pressure-temperature table which may be used for this purpose is being prepared and copies will be available shortly.

Letters to the Editor

PUBLICATION of brief reports of important discoveries in physics may be secured by addressing them to this department. The closing date for this department is five weeks prior to the date of issue. No proof will be sent to the authors. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents. Communications should not exceed 600 words in length and should be submitted in duplicate.

Single-Crystal Neutron Analysis of Tetragonal $BaTiO_3$†

B. C. FRAZER,* H. R. DANNER, AND R. PEPINSKY‡
Brookhaven National Laboratory, Upton, New York
(Received August 22, 1955)

ABOVE its ferroelectric Curie temperature $BaTiO_3$ has the cubic perovskite-type structure with Ba on the cell corners, Ti at the center, and O's on the cube faces. In the transition to the ferroelectric tetragonal phase, the atoms are freed from symmetry restrictions on their z coordinates. With Ba at the cell origin, the coordinate parameters may be written: Ba at (0,0,0); Ti at $(\frac{1}{2},\frac{1}{2},\frac{1}{2}+\delta z_{\mathrm{Ti}})$; O_I at $(\frac{1}{2},\frac{1}{2},\delta z_{\mathrm{OI}})$; O_{II} at $(\frac{1}{2},0,\frac{1}{2}+\delta z_{\mathrm{OII}})$ and $(0,\frac{1}{2},\frac{1}{2}+\delta z_{\mathrm{OII}})$. The δz displacements from cubic positions must be known accurately before theoretical understanding of ferroelectricity in $BaTiO_3$ can be achieved.

The most thorough of the several x-ray attempts on this structure were those of Kaenzig[1] and Evans.[2] The especially careful analysis of Evans finally showed that a unique x-ray solution of the problem is not possible. Because of the high Ba scattering and the small values of the δz's one is not able to separate the effects of coordinate and temperature parameters in the interpretation of diffraction data. As Evans suggested, the structure can be solved with neutrons, however. Here the scattering factors do not differ very greatly in magnitude. Moreover, they are constant with angle, and hence the range of observation may be extended without a great loss in intensity and without the problem of extending scattering factor curves beyond their known range.

A successful determination of the tetragonal $PbTiO_3$ structure was recently reported,[3] as deduced from a combined single-crystal x-ray and powder neutron diffraction study. The large tetragonal distortion of the lattice (about 6% as compared to only 1% for $BaTiO_3$) suggested at the outset that the δz's might be large enough to permit a determination by x-rays alone. After arriving at a structure that agreed very well with observed x-ray data, it was found that the agreement was not significantly affected by choosing any of the four possible combinations of algebraic signs for the displacement directions. The large cell distortion permitted sufficient resolution in a neutron powder pattern to obtain data which showed conclusive agreement with only one of the models. This study further emphasized the necessity of using neutrons on $BaTiO_3$. Because of the small cell distortion, powder methods cannot be employed. The availability of good single crystals of $BaTiO_3$ has permitted a single-crystal study, which has resulted in an unambiguous solution of the tetragonal structure.

A c-plate was cut with dimensions 12 mm×2.5 mm ×0.38 mm, with an a-axis parallel to the length. Electrodes were applied to permit unidirectional domain alignment by an electric field during the neutron observations (350 volts across the c-plate≅9200 volts/cm). Scattering data were collected on about 50 symmetry-unrelated (HOL) reflections at mean neutron wavelengths of 1.063 A and 0.905 A.

Kaenzig had reported $\delta z_{\mathrm{Ti}}=0.014$, $\delta z_{\mathrm{OI}}=-0.032$, and $\delta z_{\mathrm{OII}}=0$. Evans found two structures which gave almost equally satisfactory agreement between observed and calculated data: (1) $\delta z_{\mathrm{Ti}}=0.012$, $\delta z_{\mathrm{OI}}=-0.026$, $\delta z_{\mathrm{OII}}=0$; (2) $\delta z_{\mathrm{Ti}}=0.015$, $\delta z_{\mathrm{OI}}=-0.024$, $\delta z_{\mathrm{OII}}=-0.020$. In addition to the coordinate parameters, Evans took into account nine anisotropic temperature parameters: 2 each for Ba, Ti, and O_I, and 3 for O_{II}. He discarded the (2) structure because the temperature parameters seemed improbable.

Trial calculations were made based on the above structures and the additional models obtained by taking all combinations of signs. The correct structure appeared to lie somewhere between Evans' two structures. Trial and error calculations were made on this basis, and followed by least squares refinement. This analysis led to the following coordinate parameters: $\delta z_{\mathrm{Ti}}=0.014$, $\delta z_{\mathrm{OI}}=-0.023$, and $\delta z_{\mathrm{OII}}=-0.014$. Structure factors were calculated using these coordinates and the following isotropic temperature parameters: $B_{\mathrm{Ba}}=0.273$, $B_{\mathrm{Ti}}=0.152$, $B_{\mathrm{OI}}=0.334$, and $B_{\mathrm{OII}}=0.267$. While this structure shows excellent agreement between calculated and observed data (reliability index =0.0296) it is being checked by further measurements at high Bragg angles and by observation of (HHL) spectra from another sample.

Using the cell dimensions of Megaw,[4] $a=3.995$ A and $c=4.034$ A, the neutron results indicate a Ti shift of 0.06 A, and opposite O shifts of −0.09 A for O_I and −0.06 A for O_{II}. While at first sight these displacements are surprising when compared with those in $PbTiO_3$, since in that study the Ti and O displacements were all in the same direction, a consistent picture is obtained if one views the two structures in terms of shifts of Ti, Pb, and Ba relative to the oxygen framework, instead of fixing the origin at the Ba (or Pb) position. The shifts are then all in same direction. In $PbTiO_3$, the oxygen octahedra are deformed by the same amount as the lattice. Pb shifts 0.46 A and Ti shifts 0.30 A. In $BaTiO_3$ there is a small tetragonal distortion of the oxygen octahedra over the lattice distortion. Depending on which oxygen is "fixed," Ba shifts 0.06 A or 0.09 A, and Ti shifts 0.12 A or 0.15 A.

The $PbTiO_3$ shifts are much larger than those in $BaTiO_3$, but this might be expected from the greater lattice distortion and high polarizability of the Pb ion.

Excellent single crystals for this investigation were furnished by Dr. Kohman of the Bell Telephone Laboratories, Dr. Karan of International Business Machines in Poughkeepsie, Dr. Lefkowitz of the Glenco Corporation, and Dr. Jaffe of the Brush Development Company. The research has been supported jointly by the Brookhaven National Laboratory and contracts with the U. S. Atomic Energy Commission and the Signal Corps Engineering Laboratories.

† Under the auspices of the U. S. Atomic Energy Commission.
* Now a guest at Brookhaven from the Westinghouse Research Laboratories.
‡ Present address: Department of Physics, The Pennsylvania State University, University Park, Pennsylvania.
[1] W. Kaenzig, Helv. Phys. Acta **24**, 175–216 (1951).
[2] H. T. Evans, Jr., Technical Report No. 54, Philips Laboratories, Inc., Irvington-on-Hudson, New York, 1952 (unpublished).
[3] Shirane, Pepinsky, and Frazer, Phys. Rev. **97**, 1179 (1955).
[4] H. D. Megaw, Proc. Phys. Soc. (London) **58**, 133 (1946).

Magnetic Domains in Thin Films of Nickel-Iron*

CHARLES A. FOWLER, JR., AND EDWARD M. FRYER
Department of Physics, Pomona College, Claremont, California
(Received August 29, 1955)

CONSIDERABLE attention has been given recently to the problem of the magnetic behavior of thin ferromagnetic films.[1–5] Measurements of certain of the magnetic properties have seemed to indicate that domain configurations in general depend upon the thickness and the constants of the films in the manner predicted by Kittel,[6] namely, a single domain parallel to the film when the thickness is less than a critical value and domains oriented normal to the film surface when this value is exceeded. Direct observation of the domain structure in films has apparently been unsuccessful, although an indirect observation by optical transmission through very thin films has been reported by König.[7]

By employing the longitudinal Kerr magneto-optic method of domain observation, previously described,[8] we have succeeded in photographing the domain structure of an evaporated NiFe film (80% nickel) 5000 angstroms thick, measuring 1 in.×1¼ in. on the surface. Figure 1(A) is a composite photograph of the several portions of the complete film surface, the composite being necessary because of the rather limited photographic field of our Kerr setup. The pattern consists of a number of antiparallel, gently curving domains several millimeters wide and separated near certain of the film edges by dagger-like domains of various lengths. There is no evidence of closure domains at any of the film edges. By photographing through the microscope cover glass which serves as a backing for the evaporated

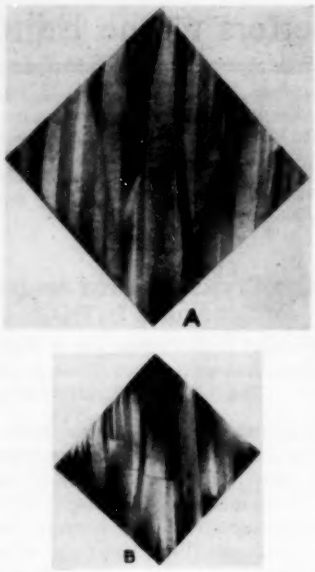


FIG. 1. Composite photographs of the domain configurations in the top (A) and the bottom (B) surfaces of a NiFe film 5000 A thick. The specimen was turned over about an axis connecting the top and bottom corners of the pictures.

alloy, the Kerr technique has revealed the domain configuration on the reverse side of the film, and this proves to be a mirror image of the front side when the film has been turned over about a longitudinal axis [see Fig. 1(B)]. Thus it follows that each domain extends completely through the film, as expected.

The domains are exceedingly sensitive to external magnetic fields. Figure 2 illustrates their behavior in a portion of the surface as the S pole of a small compass needle is brought up to the edge of the film and then withdrawn. These three photographs are part of a longer sequence which indicates that the changes caused by the small field of the perturbing pole occur by movement of the domain walls. When the pole is replaced by an unmagnetized steel needle, a smaller but still significant domain shift is observed. We have even observed changes in domain configuration to occur overnight, presumably caused by small stray fields in the laboratory.

Investigation of two other films of this material but with thicknesses of 10 000 angstroms and 20 000 angstroms, respectively, showed no evidence of surface domains. It is possible that these thicker films contain the predicted normal domains, perhaps very much smaller in surface area than those of Fig. 1. We are now attempting to photograph or otherwise detect these as a part of a systematic experimental study of the domain behavior in films of this thickness down to those of 100 angstroms or less, in a controlled uniform magnetic field.

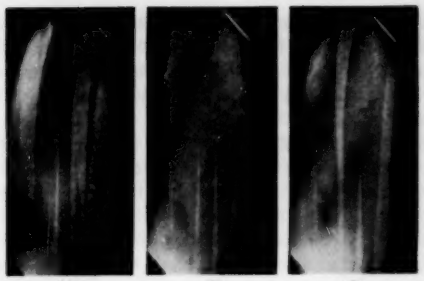


FIG. 2. Domain behavior in a portion of the film as the S pole of a small compass needle is brought up to the film edge and then removed. The needle end can be seen at the top of (B).

The authors are indebted to Dr. R. L. Conger and his group at the Naval Ordnance Laboratory, Corona, California for furnishing us with the thin film specimens and for discussions concerning their magnetic behavior.

* Supported by the Office of Naval Research.

[1] E. C. Crittenden and R. W. Hoffman, Revs. Modern Phys. **25**, 310 (1953).
[2] M. J. Klein and R. S. Smith, Phys. Rev. **81**, 378 (1953).
[3] W. Reinke, Z. Physik **137**, 169 (1954).
[4] L. E. Collins and O. S. Heavens, Phil. Mag. **45**, 283 (1954).
[5] R. L. Conger, Phys. Rev. **98**, 1752 (1955).
[6] Charles Kittel, Phys. Rev. **70**, 965 (1946).
[7] H. König, Naturwiss. **33**, 71 (1946).
[8] C. A. Fowler and E. M. Fryer, Phys. Rev. **94**, 52 (1954).

Observation of Quantum Effects in Cyclotron Resonance

R. C. FLETCHER, W. A. YAGER AND F. R. MERRITT
Bell Telephone Laboratories, Murray Hill, New Jersey

(Received August 4, 1955)

NEW cyclotron resonance lines associated with the valence band of germanium have been resolved. These are believed to be the experimental confirmation of the predictions of Luttinger and Kohn,[1] who showed that, when a crystal band edge is degenerate, a quantum mechanical treatment of the cyclotron resonances predicts a nonuniform spacing between the lower energy levels. Thus, if these levels can be preferentially populated by going to low enough temperatures, new cyclotron resonances should appear.

Measurements were made at 23.5 kMc/sec at temperatures from 4.2 to 1.3°K. For this latter temperature the Boltzmann factor, $e^{h\nu/kT}$, is 2.5, which should be large enough to provide a significant population of the lower levels. Carriers of both signs were introduced into the germanium lattice by illumination with white tungsten light. In order to study holes in the presence of much larger electron resonances, a circular cavity was used which gave circular polarization. A series of measurements was made on three slices cut from a single-crystal ingot (8×10^{12} cm^{-3} excess donors) in the three principal directions. Final alignment of the sample in the magnetic field was accomplished by rotating the sample until the appropriate electron resonances merged.

In Fig. 1 there is displayed a reproduction of the heavy-hole resonance at both 4.2°K and 1.3°K for the three directions. These reproductions represent a smoothed-out composite of a number of recorder tracings for each case. All the curves are normalized to unity at their maximum. As can be seen, new resonances appear at 1.3°K which are not present at 4.2°K in all three directions. In two instances [$m^*=0.332$ in the (111) direction and 0.262 in the (110) direction] the extra resonances were also seen at 4.2°K. The reason why these have escaped detection previously is presumably a combination of too high a microwave power (possibly raising the average electron energy to the high-energy levels) and too short mean-free-time (broad lines).

The resonances with question marks are somewhat uncertain because of proximity to other lines or because of residual electron lines arising from imperfect circular polarization. This latter also causes an appreciable effect on the shape and position of the strong line at $m^*=0.360$ in the (110) direction. The extra line in the (100) direction at $m^*=0.593$ is unmistakable at 1.3°K but vanishes at 4.2°K. No comparable line was found in the other directions.

The light-hole resonance did not display the same kind of structure. It was measured to have the values

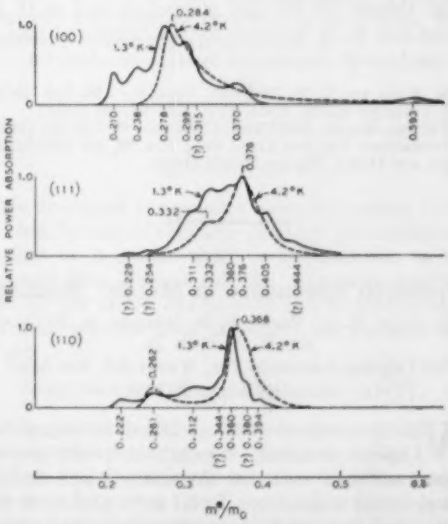


FIG. 1. Heavy-hole resonances at 4.2°K and 1.3°K in the three major crystal directions using circular polarization. The appearance of extra resonances at the lower temperature is believed to be the quantum effect.

0.0437, 0.0421, and 0.0425, in the (100), (111), and (110) directions respectively, the same at both 4.2°K and 1.3°K. Its line width was narrower than any previously reported, $\Delta H/H \doteq 0.04$ between half power points at 1.3°K, corresponding to $\tau = 3\times10^{-10}$ sec. In one direction, (100), a weak extra line was observed at $m^*/m_0 = 0.057\pm0.002$, down a factor of twenty from the main line. Although this line did not suffer a large change in intensity relative to the main resonance as the temperature was varied from 4.2° to 1.3°, it probably is also associated with quantum effects.

The ratio of the intensity of the hole resonances to the electron resonances has been found to depend critically on the sample, the temperature, the microwave power level, and the light intensity. This suggests that temporary hole traps are probably present. It is believed that these traps may have been responsible for earlier failure[2] to observe hole resonances at 1.3°K.

Because of the observed electron resonances at 1.3°K in this sample have an $\omega\tau = 80$, almost an order larger than has been previously reported,[3] it seems worthwhile reporting these measurements: $m^*/m_0 = 0.1341$ for (100); 0.0814 and 0.206 for (111); 0.0983 and 0.361 for (110). These five measurements form a consistent determination of the two constants of the energy ellipsoid: $m_l = 0.0813\pm0.002$ and $m_t = 1.600\pm0.008$. These agree with previous observations within[3] their expressed accuracy.

We are indebted to W. Kohn and J. M. Luttinger for their continual interest and stimulating discussions during the course of these measurements, to H. W. Dail for experimental assistance, to F. J. Morin and T. H. Geballe for the loan of crystals, and to H. R. Moore and W. H. Brattain for the design and loan of the mechanical chopper and synchronous detector.

[1] W. Kohn and J. M. Luttinger, Phys. Rev. **96**, 529 (1954); J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).
[2] Fletcher, Merritt, and Yager, Phys. Rev. **98**, 1560(A) (1955).
[3] Dresselhaus, Kip, and Kittel, Phys. Rev. **98**, 368 (1955); Lax, Zeiger, and Dexter, Physica **20**, 818 (1954).

Cyclotron Resonance in Metals: Bismuth

J. K. Galt, W. A. Yager, F. R. Merritt, B. B. Cetlin, and H. W. Dail, Jr.
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received August 31, 1955)

WE have observed the cyclotron resonance[1] behavior of metallic bismuth at 24 000 Mc/sec under conditions such that the steady magnetic field was along the trigonal axis. So far as we are aware, this is the first time such observations have been made successfully on a metal. The (00·1) face of a single crystal of bismuth was made to act as part of one wall of a microwave cavity, and our data consist of observations of the variation of energy absorbed in the cavity at resonance as a function of the dc magnetic field applied normal to this wall. Considerable care and effort were required to obtain clean, strain-free (00·1) surfaces of bismuth.

The experimental arrangement is diagrammed in Fig. 1. A degenerate circular cavity of coin silver in the shape of a pillbox was excited from one end through a coupling window in such a way as to produce a circularly polarized field near the axis of the cavity. The excitation of circular polarization was achieved by proper placement of the coupling window on the broad face of a rectangular wave guide, and by the adjustment of dielectric vanes inside the cavity. The bismuth sample covered a hole in the wall at the end opposite the coupling window, thus forming the cavity wall in an area where circularly polarized radiation impinges upon it. A sample of calcium copper acetate hexahydrate was in the cavity near the bismuth at all times. It was possible to establish the degree of circularity of our polarization by observing the ratio of the intensities of paramagnetic resonance in this salt when observed with the dc magnetic field in opposite directions. A ratio of at least 10 to 1 between the intensities of the two circular polarizations was maintained in taking our data.

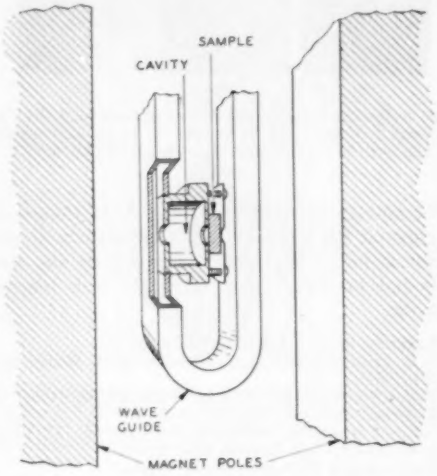


FIG. 1. Diagram of experimental geometry for cyclotron resonance experiments in bismuth. The cavity can be moved back and forth across the broad face of the wave guide so that the coupling window accepts the best approximation to circular polarization. The magnet pole faces are oriented parallel to the ends of the cavity. The sample covers a hole in one end of the cavity and thus forms part of the cavity wall at the center of this end.

The power absorption coefficient of our best sample of bismuth varied with the dc magnetic field as shown in Fig. 2 at 4.2°K. The direction of increased power absorption was established by observation of the para-

magnetic resonance in the calcium copper acetate hexahydrate sample in the cavity. The scale of H in Fig. 2 extends from 10^4 gauss in one direction to 10^4 gauss in the opposite direction. There are clearly two contributions to the absorption. One is a broad increase in absorption as the field becomes more positive, which has been dotted in the center of the figure. The other is a resonance line of more conventional shape, which however *reduces* the power absorption coefficient in the bismuth at its maximum. We associate this line with cyclotron resonance, since the increase in conductivity associated with such a resonance *increases* the Q of the cavity. We emphasize that careful determination of the shape of this line shows that it is not symmetrical about zero field. Furthermore, there is evidence that at least two lines are present (4 points of inflection). The shape of the curve in this low field region is not symmetrical, or even the sum of two simple symmetrical lines. In fact it seems to vary considerably from sample to sample, presumably because of a variation in strain.

We feel that these results can be interpreted qualitatively in terms of the plane wave analysis developed by Anderson.[2] The broad variation in the absorption coefficient is of the form expected from the interaction of the electromagnetic field and the majority carriers. We are able in bismuth to see an unusually large part of this curve because the majority carriers here have a rather small mass. The resonance line of more conventional shape we associate with a group of minority carriers which are screened from close coupling with the electromagnetic field by the majority carriers. Our best estimate from a comparison of our data at 4.2°K with Anderson's results is that the majority carriers are holes of mass approximately 0.3 m_0 and that the center of the line associated with minority carriers in Fig. 2 corresponds to holes of mass approximately $m=0.0015$ m_0, where m_0 is the mass of a free electron. These masses agree rather well with two of those quoted by Shoenberg.[3]

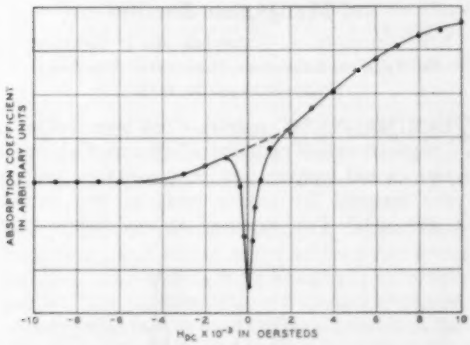


FIG. 2. Absorption coefficient *vs* dc magnetic field for circularly polarized radiation incident on the (00·1) plane of bismuth at 4.2°K. These data were taken at 24 000 Mc/sec. The vertical scale is only approximately linear. The magnetic field is normal to the (00·1) plane. Zero absorption is somewhere below the axis of abscissas.

The authors wish to thank R. E. Enz for aid in growing the bismuth crystals used in this research.

[1] J. Dorfmann, Doklady Akad. Nauk (S.S.S.R.) **81**, 765 (1951); R. B. Dingle, "Proceedings of the international conference on very low temperatures," edited by R. Bowers (Oxford, England, 1951); R. B. Dingle, Proc. Roy. Soc. (London) **A212**, 38 (1952); W. Shockley, Phys. Rev. **90**, 491 (1953); Dresselhaus, Kip, and Kittel, Phys. Rev. **92**, 827 (1953); Lax, Zeiger, Dexter, and Rosenblum, Phys. Rev. **93**, 1418 (1954); Dexter, Zeiger, and Lax, Phys. Rev. **95**, 557 (1954); Dresselhaus, Kip, and Kittel, Phys. Rev. **98**, 368 (1955).

[2] P. W. Anderson, following Letter [Phys. Rev. **100**, 749 (1955)]. See also Dresselhaus, Kip, and Kittel, Phys. Rev. **100**, 618 (1955).

[3] D. Shoenberg, Trans. Roy. Soc. (London) **A245**, No. 891, 1 (1952).

Electromagnetic Theory of Cyclotron Resonance in Metals

P. W. ANDERSON
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received August 31, 1955)

IN connection with experiments of Galt *et al.*[1] on cyclotron resonance in Bi, we have worked out the theory of the electromagnetic ("depolarizing") effects in cyclotron resonance in metals. For this purpose we confined ourselves to the simplest possible model for the electrons; that is, an isotropic effective mass and a single, classical mean free time. Under these circumstances one can define a conductivity σ of a group of electrons of mass m and mean free time τ, circularly polarized with angular frequency ω in a plane perpendicular to a steady magnetic field H:

$$\sigma=\frac{ne^2\tau}{m[1+i(\omega_c-\omega)\tau]}, \tag{1}$$

where

$$\omega_c=-eH/mc. \tag{2}$$

We considered two possible measuring setups. First, we took the case of circularly polarized plane radiation incident on a sample which is perpendicular to the magnetic field. This we call the "transverse circular" case. In this case it can be shown to be legitimate to insert (1) into the usual skin-effect loss equation and we get

$$\text{Absorption coefficient} \propto \mathrm{Re}\{[i+(\omega_c-\omega)\tau]^{\frac{1}{2}}\}. \tag{3}$$

This is plotted in Fig. 1. It can be shown that in the dissipationless metal ($\tau\to\infty$) the scale of Fig. 1 is so enlarged that the cyclotron resonance point is a transition point between a "stop" band and a "pass" band. This perhaps clarifies the peculiar shape of Fig. 1.

If, as seems to be often true in metals showing marked De Haas-van Alphen effects,[2] there is a minority group of electrons with markedly different m and τ

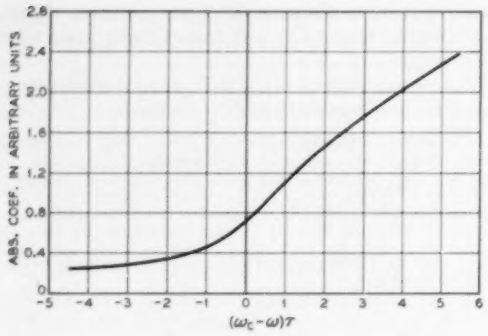


FIG. 1. Absorption coefficient for a majority group of electrons undergoing cyclotron resonance.

from the majority and much smaller total conductivity, one finds a new phenomenon: the minority electrons appear to be "screened" from strong interaction with the electromagnetic field and they will actually show a true resonance. However, under the conditions in which cyclotron resonance is observable the complex conductivity of the main group of electrons may have any phase angle. This leads to the possibility of many different line shapes for the absorption of the "screened" electrons, from a pure absorption to a pure dispersion curve. The shape depends upon where the subsidiary resonance is on the curve of Fig. 1. It should be noted that in any case there will be on the average a *reduction* in absorption coefficient due to the minority groups of electrons. Perhaps the most likely shape for these resonances is the usual line shape in metals as described by other authors.[3]

In the second experimental setup, the H-field is parallel to the sample cavity wall and one might as well use linear polarization. This we call the "longitudinal" case. Here the depolarizing effects completely eliminate the resonance of the majority group of electrons. However, the screening phenomenon can still take place, and a small group of electrons, if present, can show resonant absorption.

Results similar to Eq. (3) have recently been obtained elsewhere,[4] but the authors seemed to doubt their correctness on the basis that any depolarizing effect, if present, would appear capable of changing the result seriously. This point can be disposed of by noting how wall losses in wave-guide systems are normally calculated. That is, one usually assumes perfectly conducting walls in order to obtain the field distribution in the cavity, from which the wall currents are then obtained. This distribution can obviously contain no depolarizing fields ($\nabla\cdot\mathbf{E}=0$ everywhere). Then the finite surface impedance of the walls is introduced, assuming however that it does not appreciably perturb the field distribution, in order to compute the losses from the known currents. We feel that if the misgivings of reference 3 were sound, no calculation of cavity losses would ever have been correct. This argument, however, does not apply in the "longitudinal" case, since there currents must flow into the surface, building up space charge within the metal, which is virtually impossible.

We have also investigated approximately the limitation on the sharpness of the cyclotron resonance produced by diffusion of electrons through the skin depth. We find that the rough relationship

$$Q^2<\tfrac{1}{4}(H^2/8\pi)/nE_F \qquad (4)$$

holds: i.e., the square of the maximum possible Q is given by the ratio of magnetic field to electron kinetic energy. This ratio is probably favorable for Bi and similar poor metals,[2] giving Q's of 1–10, but will be fairly bad for normal metals except at high fields and frequencies.

Note added in proof.—F. J. Dyson has kindly called my attention to an article by M. Ia. Azbel' and M. I. Kaganov [Doklady Akad. Nauk. **95**, 41 (1954)] in which Eq. (3) is derived and the basis for a vigorous treatment of diffusion (anomalous skin effect) is laid.

[1] Galt, Yager, Merritt, Cetlin, and Dail, preceding Letter [Phys. Rev. **100**, 748 (1955)].
[2] D. Shoenberg, Trans. Roy. Soc. (London) **245**, 1 (1952).
[3] N. Bloembergen, J. Appl. Phys. **23**, 1383 (1952).
[4] Dresselhaus, Kip, and Kittel, Phys. Rev. **100**, 618 (1955).

Ferrimagnetic Resonance in Single Crystals of Manganese Ferrite

J. F. DILLON, JR., S. GESCHWIND, AND V. JACCARINO
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received August 29, 1955)

FERRIMAGNETIC resonance has been studied in single crystals of manganese ferrite over a range of frequencies and temperatures. A quantitative analysis of the material for anions leads to the formula $Mn_{0.98}Fe_{1.86}O_4$.[1] Two types of transmission cavities

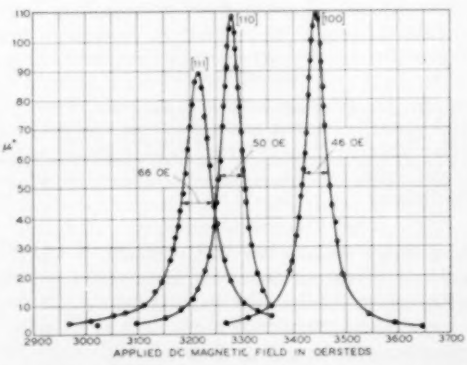


FIG. 1. Ferrimagnetic resonance line shapes measured with the dc magnetic field along the principal crystallographic directions. Measurements were made on a 0.0105-in. sphere at 9.3 kMc/sec, at room temperature.

were used; a high Q cylindrical cavity operating in the TE_{012} mode and a rectangular cavity in the TE_{102} mode. Oriented single crystal spheres of the ferrite were mounted on polystyrene rods and placed in the cavity at a position of maximum rf magnetic field which was perpendicular to the dc magnetic field. The dc magnetic field always lay in the (110) plane as the rod was rotated about its axis. The samples used were highly polished and were 0.0105 in. in diameter for 5.9 and 9.3 kMc/sec, and 0.0043 in. in diameter for the 24-kMc/sec experiments. The spheres were chosen this small to avoid dimensional resonances which result from the large value of μ' in the vicinity of resonance. Dimensional resonances were observed in a 0.014-in. sphere at X-band at room temperature such that the line width was increased 15% relative to the 0.0105-in. sphere. On the other hand, a 0.0043-in. sphere gave the same line width as the 0.0105-in. sphere within the experimental error.

Line Widths.—As the magnetic field was varied through the resonance absorption, thereby lowering the transmission through the cavity, a precision calibrated attenuator placed between the output arm of the cavity and the microwave detecting crystal was adjusted to keep the power incident on the detector at a constant level. Using the measured loaded Q of the cavity off resonance, the μ'' was then derived from the attenuator readings, thus making it independent of crystal detector characteristics. Figure 1 shows a plot of μ'' *versus* applied field at room temperature at 9.3 kMc/sec with the applied field along each of the three different symmetry directions. Line widths of 66, 50, and 46 oersteds were measured for the [111], [110], and [100] directions respectively. It is interesting to note that preliminary measurements at K-band indicate the line widths are the same, or perhaps smaller. These appear to be the narrowest ferrimagnetic resonance lines reported to date. The lines are closely Lorentz shaped over the measured region. The line width as a function of temperature in the easy direction ([111] axis) at 9.3 kMc/sec is shown in Fig. 2.

TABLE I. Anisotropy constants and g-value at several temperatures as measured on 0.0105-in. sphere at 9.3 kMc/sec. The errors in the derived quantities were estimated on the basis of reproducibility of the measurements. The values of K_1/M_s and K_2/M_s coupled with M_s as measured by M. A. Gilleo of this laboratory (M_s=351 and 493 cgs units at 300° and 77°K respectively) can be used to calculate K_1 and K_2.

	K_1/M_s	K_2/M_s	g
300 °K	− 79± 3	6±10	2.004±0.002
77 °K	−412± 5	− 5±20	2.019±0.003
4.2°K	−433±15	−200	2.060

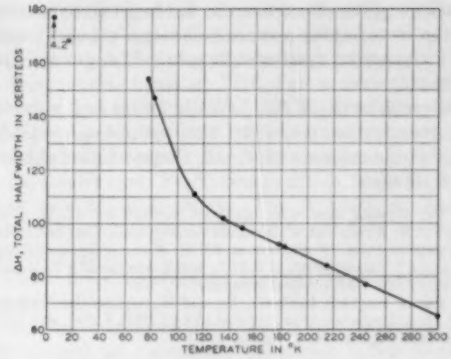


FIG. 2. Line width *vs* absolute temperature with the dc magnetic field along the easy direction ([111] axis). The line width is defined as the field interval between points of $\frac{1}{2}\mu_{max}''$.

First- and Second-Order Anisotropy Constants K_1/M_s and K_2/M_s.—By measuring the dc field required for resonance at different orientations with respect to the crystal axes the anisotropy constants can be determined as well as the g-value.[2] The temperature dependence of the anisotropy constant K_1/M_s is plotted in Fig. 3, showing an increase with decreasing temperature. The most complete temperature data to date have been taken in the region of 9.3 kMc/sec. It will be noted that room temperature points at other frequencies lie on this curve, indicating the constancy of anisotropy energy as a function of frequency.[3] The value of K_2/M_s is considerably smaller than that of K_1/M_s, and moreover is smaller than the error in its determination from 300°K down to 77°K, as indicated in Table I. It should be noted though, that at 4.2°K K_2/M_s is strikingly large.

The g-value.—Room temperature measurements at 5.9, 9.3, and 24 kMc/sec indicate that the g-value is independent of frequency, its value being 2.004±0.002. However, at 9.3 kMc/sec measurements were made over a range of temperatures and reveal a dependence, the g-value monotonically increasing with decreasing temperature reaching a value of 2.019±0.003 at 77°K. In view of the large K_2/M_s mentioned above at liquid helium temperatures, caution must be exercised in interpreting the g-value (see Table I) as well as the anisotropy constants at 4.2°K as obtained from the

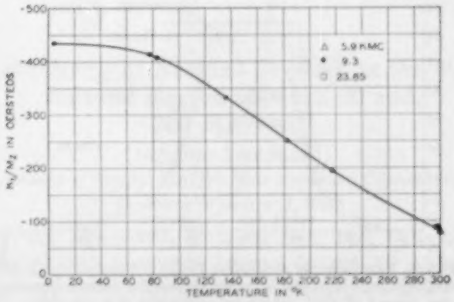


FIG. 3. Temperature variation of the first-order anisotropy constant K_1/M_s.

relations given in reference 2. A g-value so close to the free-electron g-value indicates a primarily spin-only magnetism (corresponding to a ${}^6S_{5/2}$ ground state of Mn^{++}).

We wish to thank Mr. J. White for his help in taking the measurements and we also wish to acknowledge helpful discussions with A. M. Clogston, H. Suhl, and L. R. Walker.

[1] The crystals used were obtained through Dr. G. W. Clark and Dr. R. W. Kebler of the Linde Air Products Company. Their dc resistivity was 10^4 ohm cm at 300°K. X-ray diffraction studies showed a lattice constant of 8.52 A, and it was possible to account for every line with a spinel structure.

[2] The dc magnetic fields H_1, H_2, and H_3 required for resonance at a given frequency when applied along [100], [110], and [111] directions respectively are given by

$$H_{eff}^2=\left(H_1+\frac{2K_1}{M_s}\right)^2,$$

$$H_{eff}^2=\left(H_2-\frac{2K_1}{M_s}\right)\left(H_2+\frac{K_1}{M_s}+\frac{1}{2}\frac{K_2}{M_s}\right),$$

$$H_{eff}^2=\left(H_3-\frac{4}{3}\frac{K_1}{M_s}-\frac{4}{9}\frac{K_2}{M_s}\right)^2,$$

where $h\nu=g\beta H_{eff}$. See, for example, L. R. Bickford, Jr., Phys. Rev. **78**, 449 (1950); Yager, Galt, Merritt, and Wood, Phys. Rev. **80**, 774 (1950).

[3] This result also agrees with the torque measurements of the anisotropy constant K_1 of R. M. Bozorth *et al.* [Bozorth, Cetlin, Galt, Merritt, and Yager, Phys. Rev. **99**, 1898 (1955).

Double-Frequency Magnetic Resonance in a Free Radical*

J. H. Burgess and R. E. Norberg
Department of Physics, Washington University, St. Louis, Missouri

(Received July 11, 1955)

An unusual distortion of the absorption line shape has been observed in the course of double resonance experiments on hyperfine components resolvable in aqueous solutions of the free radical ion peroxylamine disulfonate, $ON(SO_3)_2^{--}$. The experiments involve the application of intense rf radiation near resonance for one transition of the system, while detecting the resonance absorption of a second transition at low rf level.

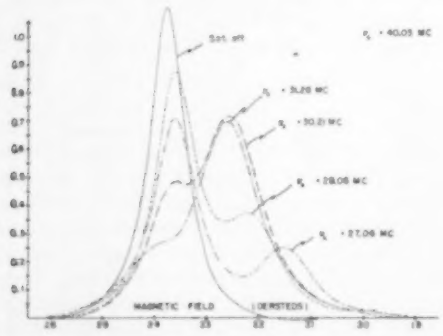


Fig. 1. Experimental absorption of the 3–4 transition (in notation of reference 1) plotted as magnetic field for various frequencies of the second rf field, which saturates the 4–5 transition. The low-amplitude detecting frequency is fixed at 40.03 Mc/sec. The saturating rf field amplitude is of the order of 2 oersteds.

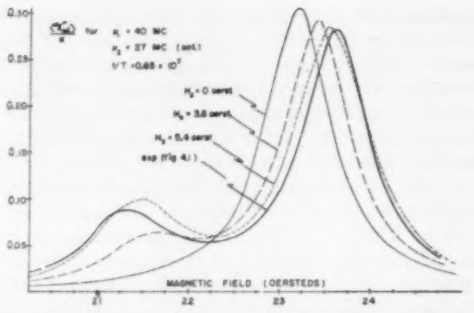


Fig. 2. Plots of Eq. 1 for three values of saturating field amplitude. $\nu_d=40$ Mc/sec and $\nu_s=27$ Mc/sec corresponding to one of the experimental curves of Fig. 1 which is the heavy curve superimposed here. The quantity T in Eq. (1) is determined from the experimental width of the unsaturated absorption curve.

If the two transitions share a common energy level, considerations of level population and lifetime predict a broadening and enhancement of the detected line as the first transition becomes saturated. However, in certain cases a "bump" is observed (see Fig. 1) on the absorption curve, at a position dependent both on the intensity of the saturating field and on the two applied frequencies.[1] The position of the bump corresponds qualitatively to the condition $\nu_s+\nu_d=\nu_{ab}+\nu_{bc}=\nu_{ac}$. ν_s, the saturating frequency, is near resonance with the transition frequency ν_{bc}, while ν_d, the detecting frequency, is near resonance with the transition frequency ν_{ab}. The experiments show that the effective gyromagnetic ratio of the combined transition, $\gamma_{eff}=d\omega_{ac}/dH$, determines the direction of the shift of the position of the bump. Increasing the amplitude of the saturating field, with frequencies ν_s and ν_d fixed, increases both the magnitude of the shift of the bump and the enhancement of the detected transition. In cases where γ_{eff} is small, the detected line is found to be broadened without observable structure.

An analysis based on the density matrix equation $d\rho/dt=(1/i\hbar)[\mathcal{H},\rho]-(1/T)(\rho-\rho_{inst})$, where ρ_{inst} is the normalized Boltzmann factor appropriate to the instantaneous total energy of the system and T is the thermal relaxation time, leads to the following expression for $\chi''(\omega_d)$, the imaginary component of the susceptibility at the detecting frequency:

$$\chi''(\omega_d)=\frac{k}{D}\left[\frac{\omega_d}{\omega_{ab}}[(x+y)^2T^2+1]-F^2\left(\frac{(x+y)}{\omega_{ab}}-1\right)\right]$$

$$+\frac{k}{D}\frac{RF^2}{1+y^2T^2+4F^2}\left[-(2x+y)\left(yT^2+\frac{1+4F^2}{\omega_{bc}}\right)\right.$$

$$\left.+(x+2y)(x+y)T^2+3(1+F^2)\frac{\omega_s}{\omega_{bc}}\right], \quad (1)$$

where $D=[x(x+y)T^2+1-F^2]^2+y^2T^2+4F^2$; $x=\omega_d-\omega_{ab}(H_0)$; $y=\omega_s-\omega_{bc}(H_0)$; $R=\omega_{bc}(H_0)/\omega_{ab}(H_0)$; $k=-|M_{xab}(H_0)|^2(\rho_a^0-\rho_b^0)$; $F=(TH_2/2\hbar)|M_{ybc}(H_0)|$; ρ^0 is the normalized Boltzmann factor appropriate to the time-independent energy of the system, and M is the magnetic moment operator. Equation (1) predicts the observed results closely if allowance is made for reasonable error in the measurement of H_s (see Fig. 2).

Although the use of ρ_{inst} is not essential to the explanation of the effects described here, its use does contribute an additional asymmetry to the line shape which should be detectable in a carefully performed experiment. A more complete description of the present work is in preparation.

* Supported in part by the Office of Naval Research and the U. S. Air Force.

[1] J. H. Burgess, Phys. Rev. **95**, 608 (1954).

Dichroism of the *F* and *M* Absorption Bands in KCl

C. Z. van Doorn and Y. Haven

Philips Research Laboratories, N. V. Philips' Gloeilampenfabrieken, Eindhoven, Netherlands

(Received August 3, 1955)

An additively colored KCl crystal, which has been quenched from high temperature and kept for some time at room temperature, shows the *M* absorption band ($\lambda=8100$ A at 77°K) in addition to the *F* band. Dichroism in the *F* and *M* bands is observed under the following conditions[1]:

The crystal (at 77°K) is strongly irradiated (beam in [100] direction) with light ($\lambda=5461$ A) polarized in the [011] direction. After a couple of minutes the *M* absorption for light polarized in the [011] direction appears to have been increased, while it has decreased in the [0$\bar{1}$1] direction of polarization. The *F* band shows just the opposite behavior (Fig. 1). After the irradiation has been cut off, the dichroism persists for at least several hours at 77°K but is destroyed after warming the crystal to about 200°K. Irradiation with light polarized in the [001] direction similarly induces a dichroism along this direction, though weaker than in the former case. Irradiation in the *M* band does not show any of these effects. The *F* band in freshly quenched crystals (no visible *M* band) does not show any dichroism.

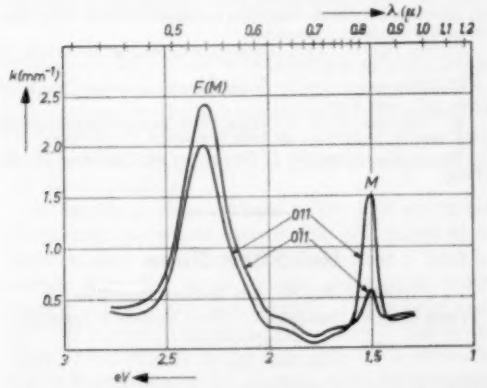


FIG. 1. Absorption of a colored KCl crystal in the [100] direction, measured with light polarized in the [011] and [0$\bar{1}$1] directions. The crystal is irradiated beforehand with light ($\lambda=5461$ A) polarized in the [011] direction. The absorption curve before irradiation is intermediate between the two shown.

From these experiments the following conclusions can be drawn:

(1) The centers responsible for the *M* absorption (*M* centers) are anisotropic; dichroism is due to non-isotropic distribution of their axes of anisotropy.

(2) The direction of these axes is along the twofold symmetry axes. (The threefold and fourfold symmetry axes can be excluded because these orientations cannot show dichroism along the [001] and [011] directions respectively.)

(3) The same conclusions as (1) and (2) hold for the centers responsible for the dichroism of the *F* absorption, and therefore these centers cannot be the *F* centers causing *F* absorption in freshly quenched crystals.

(4) Since irradiation in the *F* band causes non-isotropic distribution of the axes of anisotropy of both types of centers, there is strong evidence that these centers are identical. If that is so, irradiation in the *F* band apparently causes reorientation of the axes by ionic movement and the effect is therefore quite distinct from the dichroism found by Ueta[1] in KCl after selective bleaching of the *M* band by irradiation with polarized light in this band at room temperature.

(5) The opposite behavior of dichroism in the *F* and *M* bands shows that for one direction of orientation ([011] or [0$\bar{1}$1]) the *F* absorption is stronger and the *M* absorption weaker than for the other direction of orientation ([0$\bar{1}$1] or [011]).

[1] Weak dichroism in the *F* band was found by S. Nikitine, Compt. rend. **213**, 32 (1941); **216**, 730, 758 (1943); J. phys. radium **3**, 203 (1942). Dichroism in the *M* band was observed by M. Ueta, J. Phys. Soc. Japan **7**, 107 (1952).

Reflectivity of Several Crystals in the Far Infrared Region between 20 and 200 Microns*

Hiroshi Yoshinaga

Department of Physics, The Ohio State University, Columbus, Ohio

(Received September 2, 1955)

The reflectivities of Ge, Si, InSb, PbS, PbSe, and TlCl in the region between 20 and 200 microns were measured with the far infrared spectrograph in this laboratory.[1] Figure 1 shows the results of the measurements.

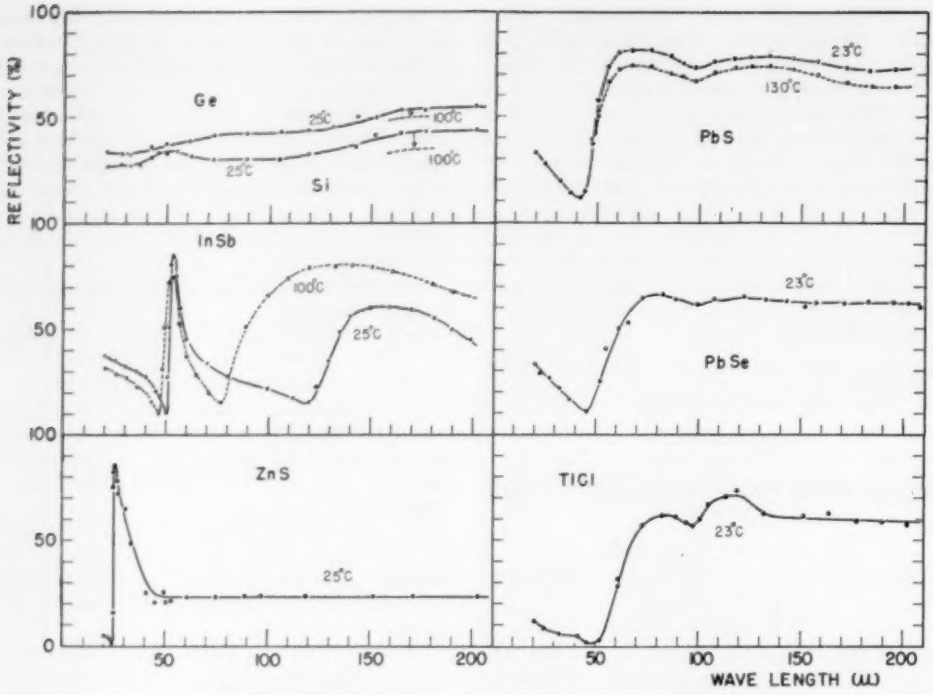


FIG. 1. Far infrared reflectivities of Ge, Si, InSb, ZnS, PbS, PbSe, and TlCl.

The reflectivities of Ge and Si are almost constant throughout the spectral region observed. For impure Ge, a peak in the 117-micron region has been reported,[2] but we were unable to find evidence of its existance with a fairly pure Ge crystal.

The reflectivity curves of InSb show a sharp maximum at 54.6 microns and a flat peak in the longer wave region. The maximum value of the peak at 54.6 microns decreases with increase of temperature, but the maximum value of the flat peak increases rapidly and the position of this peak shifts toward short wave lengths with increase of temperature. The reflectivity curve of ZnS has a sharp peak at 25.5 microns, but it is completely flat beyond 50 microns. InSb and ZnS have the same crystal form. The sharp peaks of reflectivity of InSb and ZnS appear to be due to lattice vibrations, and the flat peak of InSb may be attributed to free electrons. Details of these and other measurements on InSb are being published elsewhere.[3]

Far infrared reflectivities were previously reported by Strong[4] for PbS and by Sinton and Davis[5] for PbS and TlCl. PbS, PbSe, and TlCl have similar reflectivity characteristics, each having two flat reflectivity maxima. Measurements at 23° and 130°C show that the peak in the longer wave region does not depend on free electrons.

The author wishes to acknowledge help given by Robert A. Oetjen and Ely E. Bell. The PbSe crystal was made available by Richard Lord of the Massachusetts Institute of Technology, the InSb crystals by Leonard Eisner and T. Harman of Battelle Memorial Institute, and the ZnS crystal by Elias Burstein of the Naval Research Laboratory.

* Construction and maintenance of the far infrared spectrograph used in this investigation was supported in part by Wright Air Development Center through a contract with The Ohio State University Research Foundation.

[1] Oetjen, Haynie, Ward, Hansler, Schauwecker, and Bell, J. Opt. Soc. Am. **42**, 559 (1952).

[2] H. Y. Fan and M. Becker, *Semiconducting Materials* (Academic Press, Inc., New York, 1951).

[3] H. Yoshinaga and R. A. Oetjen, Phys. Rev. (to be published).

[4] J. Strong, Phys. Rev. **38**, 1818 (1931).

[5] W. M. Sinton and W. C. Davis, J. Opt. Soc. Am. **44**, 503 (1954).

Hall Effect Noise

JAMES J. BROPHY AND NORMAN ROSTOKER
Armour Research Foundation of Illinois Institute of Technology, Chicago, Illinois

(Received July 29, 1955)

EXCESS noise, characterized by a noise power density with a $1/f$ spectrum, is the major form of electrical noise in semiconductors, particularly at low frequencies. It may be present in single-crystal semi-

conductor samples with "noiseless" electrodes and no obvious external discontinuities.[1] It has been suggested that the effect is caused by conductivity modulation of the semiconductor due to fluctuations in carrier density. We have undertaken to examine variations in carrier density in semiconductors exhibiting excess noise, through the use of the Hall effect.

Specimens used for this study are *n*-type single-crystal germanium, of approximately 10 ohm-centimeter resistivity. They were sandblasted into "bridges" similar to those used by Montgomery.[1] This shape is known to be most suitable for noise investigations as it eliminates effects of metallic probe electrodes in the current path. To increase the noise level of some samples, they were plastically deformed by bending in air at 600°C.

The excess noise observed in these specimens is found to have the usual $1/f$ spectrum of excess noise in the frequency interval studied, 1 cps to 10 kc/sec.

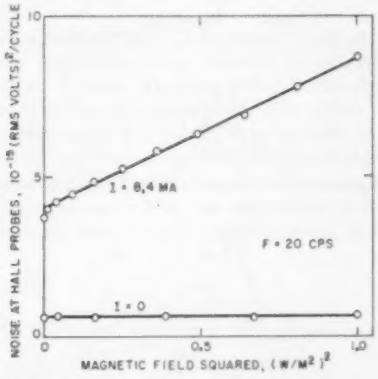


FIG. 1. Noise observed at the Hall probes as a function of magnetic field squared. The curve for $I=0$ represents amplifier noise. The intercept of the $I=8.4$ ma curve at $H=0$ is "unbalanced" noise. The linear variation of noise with H^2 should be noted.

A magnetic field of 1.0 weber/m^2 perpendicular to the plane of the sample and to the direction of current flow had a negligible effect on the magnitude and spectrum of the noise. It has been demonstrated experimentally that the noise was generated in the germanium and was not due to contacts.

To determine noise voltages at the Hall electrodes, the sample was placed between the pole pieces of an electromagnet supplied with current from a bank of storage cells. The noise voltages observed at a frequency of 20 cps are plotted in Fig. 1. The noise level for $I=0$ represents amplifier and thermal noise in the input circuit and illustrates that no noise is generated due to fluctuations in the magnetic field. With current passing through the sample, noise is observed at the probes at zero magnetic field. This is analogous to the usual "unbalance" voltage encountered in the Hall effect.

Additional noise is observed when the magnetic field is applied and the mean square noise voltage increases linearly with the square of the field, as shown in Fig. 1. This behavior is as expected in the Hall effect, and also shows that there is little correlation between the unbalance excess noise voltage and the Hall noise voltage, for if correlation existed the observed noise would have, in addition, a linear dependence upon field.

Residual current fluctuations may be expected also to contribute to noise at the Hall probes and this component would be difficult to distinguish from the Hall noise. We compute residual current fluctuations to be $\langle \Delta I^2 \rangle = 5\times10^{-21}$ amp^2/cycle under the conditions pertaining to Fig. 1. From the known Hall constant this contribution to noise at the Hall probes is of the order of 2×10^{-19} volt^2/cycle, much below the level observed. We conclude that the field-dependent noise voltage observed is truly Hall noise and may be interpreted in terms of carrier density fluctuations. The excess noise and the Hall noise may be represented by the following expressions in the frequency interval 1 cps to 10 kc/sec:

$$\langle \Delta V_I^2 \rangle = 7.3\times10^{-16} I^{2.3} f^{-1.0} \text{ volt}^2/\text{cycle},$$
$$\langle \Delta V_H^2 \rangle = 7.4\times10^{-16} I^{2.3} f^{-1.0} H^{2.0} \text{ volt}^2/\text{cycle},$$

where I is the sample current in milliamperes, f is the frequency, and H is the magnetic field in webers/m^2.

These noise voltages may be interpreted in terms of conductivity modulations caused by fluctuation in carrier density in the following manner. With zero magnetic field the potential is simply that due to the current, while the additional voltage with magnetic field is given by the usual Hall relation. The two potentials are thus:

$$V_I = RI \quad \text{and} \quad V_H = R_H IH/t, \tag{1}$$

where R is the sample resistance between probes, R_H is the Hall constant, and t is the sample thickness. If we ascribe fluctuations in these potentials to variations in

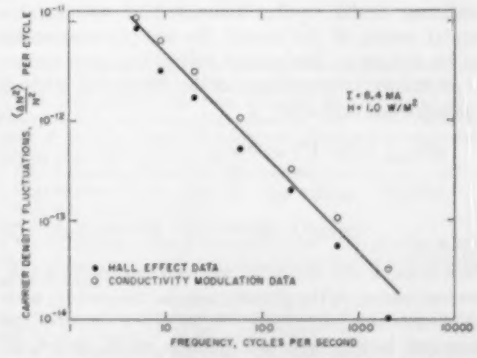


FIG. 2. Spectral density of carrier density fluctuations as a function of frequency as determined from excess noise measurements (open circles) and Hall noise (solid circles). The inverse frequency spectrum is clearly illustrated.

R and R_H caused by fluctuations in carrier density, n, it may be shown that on the basis of a simple, single-carrier model:

$$\langle \Delta V_I^2 \rangle = V_I^2(\langle \Delta n^2 \rangle / n^2) \quad \text{and}$$
$$\langle \Delta V_H^2 \rangle = V_H^2(\langle \Delta n^2 \rangle / n^2), \quad (2)$$

where V_I is the dc potential drop between the probes and V_H is the usual Hall voltage. It is convenient to treat the data in this manner, since all potentials are directly measurable and the results are independent of sample dimensions and calibration of the magnetic field. The quantities V_I and V_H were determined for the sample by standard techniques.

Carrier fluctuations calculated from Eqs. (2) are shown in Fig. 2. The agreement between the results is within the experimental errors of the present data. In particular, the carrier density fluctuations as directly measured by the Hall effect have the $1/f$ spectrum and the proper order of magnitude to account for excess noise. We believe that these results constitute the first proof that excess noise is a result of carrier density fluctuation.

[1] H. C. Montgomery, Bell System Tech. J. **31**, 950 (1952).

Scattering of Neutrons by Phonons in an Aluminum Single Crystal

B. N. Brockhouse and A. T. Stewart
Physics Division, Atomic Energy of Canada, Limited, Chalk River, Ontario, Canada
(Received August 29, 1955)

IT has long been realized[1] that the normal vibrational modes of a crystal could be determined from measurements of the energy distribution of coherently scattered neutrons. The basic theory of inelastic coherent scattering is largely due to Weinstock.[2] The neutron gains or loses energy corresponding to the annihilation or production of one or more phonons in the crystal. The zero-phonon process represents Bragg scattering. If this process is avoided by selecting the angular setting of the crystal, the one-phonon process can be dominant. The partial differential cross section for production of one phonon of the fth normal mode in a monoatomic crystal is[3]

$$G_1^f = \frac{\pi^2 \hbar \sigma_{\text{coh}}}{2MBN} \cdot \frac{[(\mathbf{k}-\mathbf{k}')\cdot\boldsymbol{\alpha}_f]^2}{\omega_f(1-x_f)} \frac{k'}{k} \times \exp(-2W_f^r)\cdot\delta(\mathbf{k}-\mathbf{k}'-2\pi\boldsymbol{\tau}+\mathbf{q}), \quad (1)$$

where $\mathbf{k}$ and $\mathbf{k}'$ are the initial and final neutron propagation vectors, ω_f is the phonon angular frequency, $\boldsymbol{\alpha}_f$ is the phonon unit polarization vector, $\boldsymbol{\tau}$ is a vector of the reciprocal lattice, $\mathbf{q}$ is the phonon wave vector, $x_f = \exp(-\hbar\omega_f/k_BT)$, N is the number of atoms in the crystal, B is the volume per atom, and M is the mass of the atom, and $\exp(-2W_f^r)$ is the Debye-Waller factor. The δ function in this expression represents the momentum conservation between neutron, phonon and crystal. In addition, the neutron-phonon system must conserve energy so that[4]

$$E' = \frac{\hbar^2 k'^2}{2m} = E - \hbar\omega_f = \frac{\hbar^2 k^2}{2m} - \hbar\omega_f. \quad (2)$$

When these energy and momentum conditions are simultaneously satisfied, neutrons are scattered with an intensity given by (1).

Since $\mathbf{q}$ is restricted to the zone around each reciprocal lattice point $\boldsymbol{\tau}$, this latter can be unambiguously assigned. Then the two conservation equations are sufficient to determine the angular frequency ω in terms of the phonon wave vector $\mathbf{q}$. For a monoatomic lattice $\omega(q)$ is expected to separate into three branches corresponding in the low-frequency limit to the longitudinal and two transverse polarizations of sound waves. In principle the directions of the polarization vectors, $\boldsymbol{\alpha}$, can be obtained by repeated observation of the same point in a zone of reciprocal space using different initial and final neutron momenta, and comparison of the intensities of the observed neutron groups with Eq. (1).

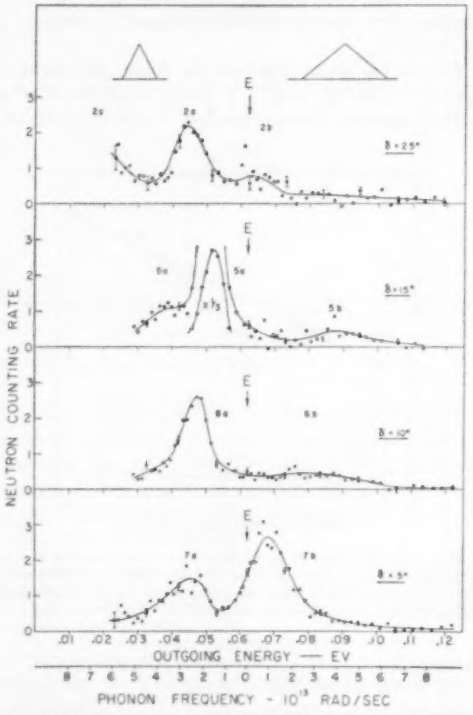


FIG. 1. Typical energy distributions of neutrons inelastically scattered by an aluminum single crystal and approximate resolution functions. The incident neutron energy E is indicated by the arrows.

Monoenergetic neutrons provided by a crystal spectrometer (wavelength 1.14_8 A, energy 0.062_2 ev) were scattered at an angle of 95.1° by an aluminum single crystal oriented so that its (1,1,1) axis was in the plane of the spectrometer and its $(0,1,\bar{1})$ axis vertical. The energy distribution of the scattered neutrons was obtained by examination with another crystal spectrometer.[5] The angle and wavelength used satisfy the Bragg condition for the (3,3,3) reflection when the crystal is oriented with its $(\bar{1},\bar{1},\bar{1})$ direction bisecting the angle of scattering. Measurements were made with the crystal rotated about the vertical $(0,1,\bar{1})$ direction by an angle δ from the (3,3,3) Bragg scattering position, keeping the angle of scattering constant. Energy distributions have been obtained for eleven angular settings of the crystal, all at room temperature, of which a selection is shown in Fig. 1. Neutron groups are seen which represent both energy losses and gains, corresponding to one-phonon production and annihilation. The estimated resolution function is shown for two energies.

Figure 2 shows the $(0,1,\bar{1})$ plane of the reciprocal lattice of aluminum, each reciprocal lattice point surrounded by its zone. The heavy lines in the figure are used to illustrate the experiment. The initial neutron propagation vector $\mathbf{k}$ is defined in magnitude and direction by the incident energy and the orientation of the specimen crystal. The direction of the outgoing neutron of propagation vector $\mathbf{k}'$ is then determined by the angle of scattering, fixed in these experiments at 95.1°. The

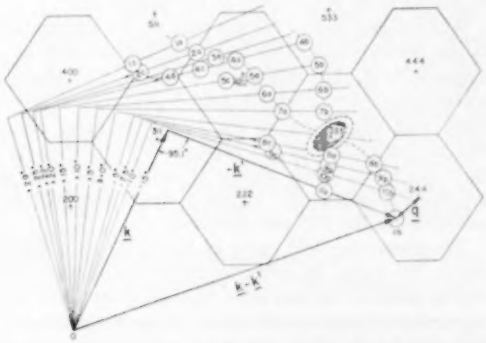


FIG. 2. Reciprocal lattice diagram of aluminum showing positions of scattered neutron groups, and approximate resolution function in reciprocal space. The position of 2*c*, 4*d*, and existence of 4*b*, 5*b*, 8*c*, and 9*c* is somewhat doubtful. One group (2*b*) is believed to be spurious and has not been entered.

magnitude of $\mathbf{k}'$ is found from the measured energy of the neutron group. The phonon vector $\mathbf{q}$ is then determined by completing $\mathbf{k}-\mathbf{k}'$ to the nearest reciprocal lattice point. The angular frequency of the phonon is obtained by Eq. (2), an auxiliary frequency scale being given in Fig. 1. An estimate of the resolution in reciprocal space is shown as the shaded area in Fig. 2. Labels have been attached to reciprocal space points corresponding to the neutron groups in Fig. 1.

In Fig. 3. the results have been plotted to show the relation between the angular frequency and wave vector of the observed phonons. This relation is a sinusoidal function for a one-dimensional lattice and

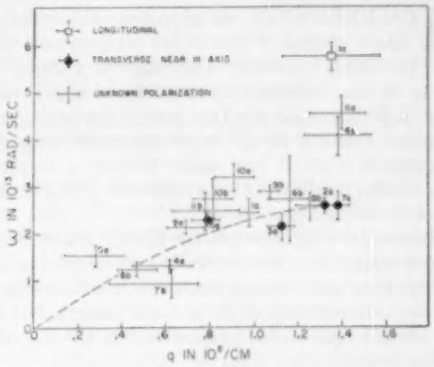


FIG. 3. Relation between ω and q for observed phonon groups, with estimated errors.

also for the transverse branches in the {1,1,1} directions of a face-centered cubic lattice.[6] The dashed curve is the sine function which has a slope at low frequencies corresponding to a velocity of sound of 3080 m/sec for transverse modes. Of the observed phonons, those which may be thought of as "transverse"[7] and having directions near the {1,1,1} axes are indicated by the black circles. Other phonons seem to be of higher frequencies. This is in accord with the results of Rosenstock[8] who showed that the (1,1,1) transverse waves have the lowest limiting frequency and that waves with other polarization or direction have limiting frequencies ranging up to about twice this value.

The experiments will be continued using this method as well as a complementary time-of-flight[9] technique.

We wish to thank Mr. G. R. DeMille for his assistance in analyzing the data and Dr. D. G. Hurst for helpful discussions.

[1] J. M. Cassels, *Progress in Nuclear Physics* (Academic Press, Inc., New York, 1950), p. 214; D. A. Kleinman, Phys. Rev. **86**, 622 (1952). The problem has been discussed in some detail by G. Placzek and L. Van Hove, Phys. Rev. **93**, 1207 (1954). An experiment of the type described below was attempted unsuccessfully by one of us (B. N. B.) in 1952.

[2] R. Weinstock, Phys. Rev. **65**, 1 (1944).

[3] See reference 2, Eq. (36).

[4] See reference 2, Eq. (39).

[5] B. N. Brockhouse, Phys. Rev. **99**, 601 (1955).

[6] G. L. Squires (private communication). This relation is valid for a crystal model with first and second nearest neighbor harmonic forces.

[7] In the low-frequency limit the polarization of the phonon can be assigned as longitudinal or transverse if the angle between $\mathbf{q}$ and $\mathbf{k}-\mathbf{k}'$ is nearly 0° or 90° respectively.

[8] H. B. Rosenstock, Phys. Rev. **97**, 290 (1955).

[9] Similar experiments using this technique are being performed by B. Jacrot at Saclay [Compt. rend. **240**, 745 (1955)] and by Carter, Hughes, Palevsky, and Zimmerman at Brookhaven [Phys. Rev. **99**, 611(A) (1955)].

Ultrasonic Attenuation in Superconducting and Normal-Conducting Tin at Low Temperatures

H. E. BÖMMEL
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received Aufust 23, 1955)

MEASUREMENTS on ultrasonic absorption in single crystals of lead at low temperatures have been previously reported.[1] They showed a strong increase in the temperature range from about 10°K -1.5°K if the sample was kept normal-conducting, and a sudden decrease at the superconducting transition temperature when it was cooled without a magnetic field. Similar results have been reported by Mackinnon[2] in polycrystalline tin.

Theoretical considerations by Kittel,[3] Mason,[4] and Morse[5] suggest that, at sufficiently high frequencies and low temperatures, the main ultrasonic attenuation is due to the interactions of the lattice vibrations with the free electron gas and is proportional to the electrical conductivity.

New measurements were carried out in single crystals of tin of two different purities: 99.98% and 99.998%. The results in the less pure tin were similar to those in lead and have already been reported.[4]

For the measurements in the purer tin, two cylindrical samples 21 mm in diameter, 16.5 mm and 9 mm long, respectively, were used. Special care was taken not to disturb the crystalline structure while preparing the specimens. They were cut with an acid-string saw and ground in liquid nitrogen.

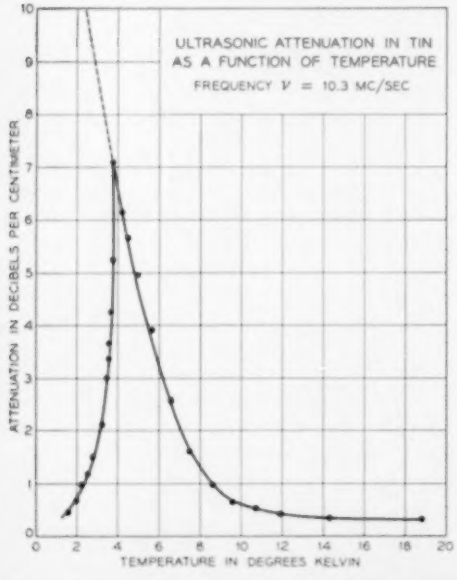


FIG. 1. Ultrasonic attenuation in a single crystal of tin as a function of temperature. Frequency $\nu = 10.3$ Mc/sec. The dotted line indicates the attenuation in the normal conducting state. Its value at 1.6°K is about 12 db/cm.

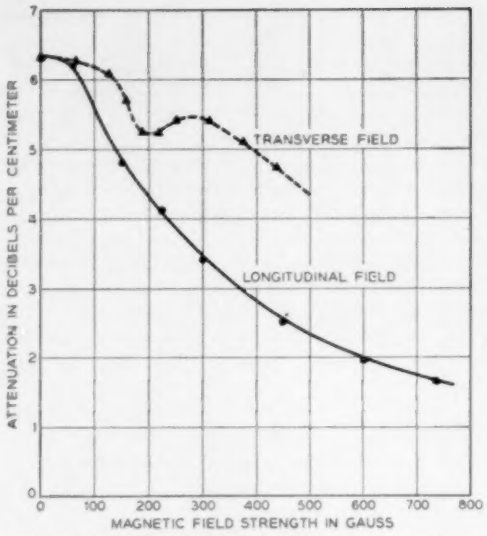


FIG. 2. Ultrasonic attenuation in a single crystal of tin as a function of magnetic field strength. Frequency $\nu = 10.3$ Mc/sec. Temperature $T = 4.2$°K.

Figure 1 shows the resulting attenuation *versus* temperature curve, the maximum attenuation being about 100 times larger than in the less pure tin at the same frequency. From this, the electrical conductivity can be estimated to be of the order of 10^{10} ohms^{-1} cm^{-1}.

While in lead and in the less pure tin the attenuation in the normal-conducting state was independent of a magnetic field, the attenuation in the pure tin showed a very pronounced magnetic field dependence. It seems likely that this is due to the magnetoresistive effect. One can estimate that at a conductivity of 10^{10} ohms^{-1} cm^{-1} this effect should appear at fields of the order of several hundred gauss. Figure 2 shows the field dependence for the 16.5-mm specimen.

The transverse field curve of Fig. 2 shows a dip at about 200 gauss. In order to study this further, the shorter sample was used since, because of its lower loss, it permitted measurements up to higher frequencies. In this sample the effect was more pronounced and appeared in the longitudinal as well as in the transverse field. Measurements were made at 10.3, 20, and 28 Mc/sec. Within the accuracy of the experiments the position of the dip shifted to higher fields proportional to the frequency.

No definite explanation can yet be given for the above-mentioned phenomenon. The electrical conductivity and therefore the mean free path of the electrons are so high that one may have to take into account size effects as they occur when the mean free path becomes comparable to the dimensions of the specimen. In this case special magnetoresistive effects

due to boundary scattering[6] should occur, and one might suggest that the observed phenomenon is due to a superposition of the ordinary magnetoresistive effect with those mentioned above. A more systematic investigation, especially the dependence on size, frequency, and crystal orientation, is in progress.

The author wishes to thank Mr. G. E. Devlin for the preparation of the pure tin crystal, and Mr. T. B. Bateman for help with the experiments.

[1] H. E. Bömmel, Phys. Rev. **96**, 220 (1954).
[2] L. Mackinnon, Phys. Rev. **98**, 1181(A) (1955).
[3] C. Kittel, Phys. Rev. **98**, 1181(A) (1955).
[4] W. P. Mason, Phys. Rev. **97**, 557 (1955).
[5] R. W. Morse, Phys. Rev. **97**, 1716 (1955).
[6] See for instance E. H. Sondheimer, Phys. Rev. **80**, 401 (1950).

Nuclear Spin-Lattice Relaxation in Solid Hydrogen

TADASHI SUGAWARA
The Research Institute for Iron, Steel, and Other Metals, Tôhoku University, Sendai, Japan
(Received August 22, 1955)

IN solid hydrogen, the crystalline potential is not the same for different *ortho*-molecules and in general has no particular symmetry, as the solid consists of a random mixture of *ortho*- and *para*-molecules. Accordingly, the lowest rotational state ($J=1$) of the *ortho*-molecule splits into three sublevels, whose separations are expressed in terms of the potential. Most of the linewidth data of the nuclear resonance absorption[1,2] were well accounted for by the above supposition originally made by Reif and Purcell. At temperatures not too low, the transition among the rotational sublevels can take place with large probability and this will give the dominant mechanism for the nuclear spin-lattice relaxation. The spin-lattice relaxation time is expected to be distributed over a wide range, as the transition probability can be related to the crystalline potential. The main purpose of this note is to get information on the distribution of the relaxation time from the saturation experiments.

All the measurements were carried out at 8.2 Mc/sec by using an rf spectrometer of the Pound-Watkins type.

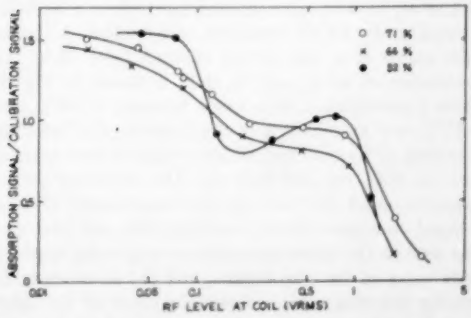


FIG. 1. The saturation curve of the central peak at 4.2°K in solid hydrogen containing 71, 66, and 52% *ortho*-hydrogen.

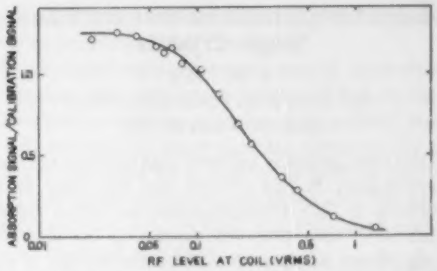


FIG. 2. The saturation curve of the side peaks at 1.2°K in solid hydrogen containing 69% *ortho*-hydrogen.

Derivatives of absorption lines at different oscillation levels were obtained in samples containing various concentrations of *ortho*-hydrogen at temperatures below 4.2°K. As was reported previously,[2] the spectrum in high-concentration hydrogen was a single peak with a width of several gauss at high temperatures; as the temperature is reduced, this peak became flanked by two side peaks of about 40-gauss separation. The line width of the central peak decreased with increasing rf level, while the separation of the side peaks remained constant. Some of the saturation curves are shown in Fig. 1 and Fig. 2 by plotting the ratio of the maximum deflection of the recording meter to the calibration signal as a function of the rf voltage at the sample coil. The curve for the central peak, which does not follow the form usually observed in systems having a single relaxation time,[3] shows that the relaxation time is not unique as had been expected. A theoretical treatment of the saturation phenomena in solid hydrogen would be necessary in order to obtain the distribution function of the relaxation time. However, we may be certain that the distribution has a nearly flat maximum in high-concentration hydrogen and two peaks in low-concentration hydrogen. A single relaxation time is expected from the form of the saturation curve of the side peaks. A preliminary value of about 0.5 second has been obtained from this curve, and this value is in disagreement with the result of Hatton and Rollin.[4]

The effect of paramagnetic oxygen upon the relaxation was also studied in connection with its remarkable effect on the *ortho-para* conversion rate in solid hydrogen.[5] Oxygen up to about 0.05% did not show any observable effect on the saturation curve. An order-of-magnitude estimation of the relaxation time due to the magnetic interaction between proton and oxygen (0.05%) gives a fairly large value, which seems to give no measurable effect on the saturation curve under the experimental conditions of this research.

A full account of this work will be published elsewhere.

[1] F. Reif and E. M. Purcell, Phys. Rev. **91**, 631 (1953).
[2] Sugawara, Masuda, Kanda, and Kanda, Science Repts. Research Insts. Tôhoku Univ. **A7**, 67 (1955).
[3] Bloembergen, Purcell, and Pound, Phys. Rev. **73**, 679 (1948).
[4] J. Hatton and B. V. Rollin, Proc. Roy. Soc. (London) **A199**, 222 (1949).
[5] H. Kobayashi and E. Kanda (to be published).

Edge Electroluminescence from ZnS Single Crystals

R. W. Smith
RCA Laboratories, Princeton, New Jersey
(Received August 18, 1955)

It has been generally believed that holes do not move appreciably in ZnS crystals and that they do not contribute to ZnS electroluminescence (EL). Several models have been proposed expressing the view that the significant EL is generated by the motion of electrons in an electric field of intensity near the dielectric breakdown strength.[1,2] The purpose of this letter is to present data indicating that both free electrons and holes may contribute to the EL from ZnS.

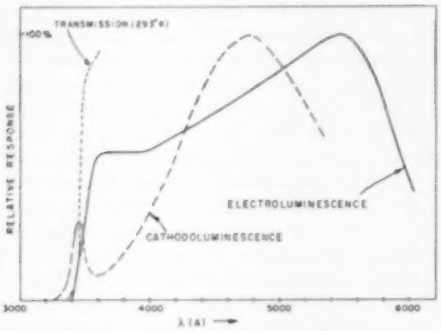


Fig. 1. Spectral emission curves for electroluminescence from ZnS single crystal and cathodoluminescence from ultrapure ZnS powder. The EL curve was recorded from 3000 to 6000 A with a resolution 100 A.

Figure 1 shows the ultraviolet and visible electroluminescence typical of several ZnS crystals measured. This dc electroluminescence was obtained with Ga electrodes placed on single crystals prepared by two separate groups: (1) crystals grown by W. W. Piper of the General Electric Company, and (2) crystals grown by E. L. Lind of these Laboratories. Impurities were not intentionally added to these crystals. The extent to which EL is observed in a crystal varies considerably with different crystals. With the above crystals, the electrodes were placed on opposite faces of the crystal, and separated somewhat more than the crystal thickness. The EL observed here is not confined to a particular electrode and there is reasonable evidence that it comes from the crystal between the electrodes. This is in agreement with the extended EL reported from ZnS by Frankl.[3] The fact that the EL is not confined to the electrodes but is observed in the bulk of the crystal is consistent with the view that electronic transitions, giving light emission, are occurring in the volume of the insulator.

The point to be emphasized in this letter is that electroluminescence is obtained in the region of the absorption edge in addition to the visible emission. The room temperature location of the ZnS absorption edge (3400 A ≈3.6 ev) is indicated by Van Doorn's[4] transmission curve for a flat plate of ZnS~0.2 mm thick, and by Shrader's cathodoluminescence[5] curve. The observation of edge emission is evidence for the recombination of free electrons and holes.

Similar observations have been reported[6] for CdS and it is not unreasonable to assume that similar mechanisms operate in both crystals. A model in which electrons and holes are injected, from opposite electrodes, into the insulator and subsequently recombine, directly to give edge emission, and through centers for the remaining emission, is consistent with these observations. This mechanism is to be contrasted with excitation by electrons in a high field at a cathode barrier, proposed by Piper and Williams[2] and with the stacked junctions proposed by Loebner[7] and Frankl.[3]

We want to thank Dr. Piper and Dr. Lind for the ZnS crystals and Dr. R. E. Shrader for permission to use his unpublished cathodoluminescence curve.

[1] D. Curie, J. phys. radium **13**, 317 (1952).
[2] W. W. Piper and F. E. Williams, Phys. Rev. **87**, 151 (1952); Brit. J. Appl. Phys. Supplement No. 4 (1955).
[3] D. R. Frankl, reported at Spring Meeting of The Electrochemical Society, Cincinnati, Ohio, May, 1955 (unpublished).
[4] C. Z. Van Doorn, Physica **20**, 1155 (1954).
[5] In this case ultrapure ZnS powder prepared without fluxing agents was used. This powder does not luminesce under uv excitation.
[6] R. W. Smith, Phys. Rev. **93**, 347 (1954); Phys. Rev. **98**, 1169(A) (1955).
[7] E. E. Loebner and H. Frund, Phys. Rev. **98**, 1545(A) (1955).

Thermoluminescence Measurements of Electroluminescent ZnS:Mn Films

R. E. Freund
Mellon Institute, Pittsburgh, Pennsylvania*
(Received August 22, 1955)

Thermoluminescence measurements were made on films of chemically deposited ZnS:Mn formed by the vapor reaction method described by Studer, Cusano, and Young.[1] A number of films, in addition to being electroluminescent upon the application of an ac field, also emitted light when a dc field, approximately 3×10^5 volts/cm, was applied. A typical glow curve of a film which emitted light upon the application of an ac and dc field is shown in Fig. 1. These measurements were made between −196°C and +25°C, and calculations of the three depths based on a formula of Randall and Wilkens[2] yielded the values of 0.37 ev, 0.41 ev, and 0.43 ev. The following points should be noted: (a) the formula of Randall and Wilkens is based on a glow curve containing only one peak and was used in the above calculations to give an approximate value of the trap depths, and (b) an exponential heating rate was used to obtain the curve of the figure since linear heating rates yielded values which deviated

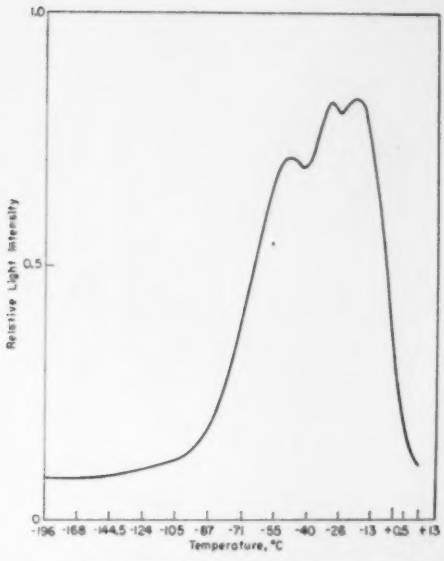


FIG. 1. Glow curve with an exponential heating rate of chemically deposited ZnS:Mn.

by a maximum of only 3% from the glow peaks illustrated in the figure.

Halsted and Koller have reported[3] thermoluminescence measurements of films prepared in a similar manner as having only one glow peak corresponding to an 0.3-ev trap. Their films had a resistivity of 10^{13} ohm-cm as compared to a resistivity of 10^{16} ohm-cm for films prepared in this laboratory.

The voltage-brightness relationship of these films at 60 cps has been previously described,[4] but no similar measurements were made on the light emitted when a dc field was applied across the film. Although the light emission continued for a number of hours it was very faint, and it also should be mentioned that the ability to reproduce consecutive films that responded similarly to a dc field was very poor.

The author does not feel justified in calling this light emission dc electroluminescence. It may possibly be due to random local field breakdowns across imperfections throughout the film, resulting in the surrounding crystal cells seeing a pulsating dc field. Inasmuch as Halsted and Koller observed no light emission when a dc potential was placed across films with a resistivity of approximately 10^{13} ohm-cm (except for a transient flash during the application and removal of the dc field), while in the present work the application of a dc field did produce luminescence in films with a resistivity of 10^{15} ohm-cm, it is possible a relationship exists between the film resistivity and light emission when a dc field is placed across the film.

* Multiple Fellowship on Computer Components sustained by the Electronics Directorate of the U. S. Air Force Cambridge Research Center, Research and Development Command.

[1] Studer, Cusano, and Young, J. Opt. Soc. Am. **41**, 559 (1951).
[2] J. T. Randall and M. H. F. Wilkens, Proc. Roy. Soc. (London) **A184**, 373 (1945).
[3] R. E. Halsted and L. R. Koller, Phys. Rev. **93**, 349 (1954).
[4] F. A. Schwertz and R. E. Freund, Phys. Rev. **98**, 1134 (1955).

Letters to the Editor:

Single-Crystal Neutron Analysis of Ferromagnetic [illegible]
. . . . . R. J. Harvey, and S. Siegel 745

Magnetic Domains in Thin Films of Nickel-Iron
. . . . . [illegible], Jr., and William M. Fiske 745

Observation of Quantum Effects on Cyclotron Resonance
. . . . . W. A. Yager, and J. K. Merritt 747

Cyclotron Resonance in Metallic Bismuth
. . . . . J. K. Galt, W. A. Yager, F. R. [illegible], B. B. Cetlin, and H. W. Dail, Jr. 748

Thermodynamic Theory of Cyclotron Resonance in Metals . . . . . P. W. Anderson 749

Paramagnetic Resonance in Single Crystals of [illegible] Perrite
. . . . . J. F. [illegible], Jr., S. Geschwind, and V. Jaccarino 750

Double-Frequency Magnetic Resonance in a Free Radical . . . . . J. H. Burgess and R. E. Norberg 752

Resolution of the F and M Absorption Bands in KCl . . . . . C. Z. van Doorn and Y. Haven 753

Ferromagnetism of General Crystals in the Far Infrared Region between 20 and 200 Microns
. . . . . Hiroshi Yoshinaga 754

Half-Metal Films . . . . . James J. Brophy and Norman Rostoker 755

Scattering of Neutrons by Phonons in an Absorbing Single Crystal
. . . . . B. N. Brockhouse and A. T. Stewart 756

Ultrasonic Attenuation in Superconducting and Normal Tin Containing Tin at Low Temperatures
. . . . . R. E. Dransfeld 758

Acoustic Spin-Lattice Relaxation in Gallium Ferrite . . . . . Tadaoki Sagawara 759

Ferromagnetic Resonance from Thin Single Crystals . . . . . R. W. Smith 760

Ferromagnetic Resonance Measurements of Permeability in Nickel Films . . . R. E. Fremont 760